

Systematische Übersichtsarbeiten und Meta-Analysen

—

Anwendungsbeispiele und empirisch-methodische Untersuchungen

HABILITATIONSSCHRIFT

zur Erlangung der Lehrbefähigung
für das Fach
Epidemiologie

vorgelegt dem Fakultätsrat der Medizinischen Fakultät Charité
der Humboldt-Universität zu Berlin

von
Dr. Klaus Linde
geboren am 16. Juli 1960 in München

Präsident: Prof. Dr. rer. nat. J. Mlynek

Dekan: Prof. Dr. Joachim W. Dudenhausen

Gutachter: 1. Prof. Dr. J. Kleijnen

2. Prof. Dr. K. H. Jöckel

Datum des öffentlich-wissenschaftlichen Vortrages: 03.12.2002

Eidesstattliche Versicherung

gemäß Habilitationsordnung der Charité

Hiermit erkläre ich, dass

- keine staatsanwaltschaftlichen Ermittlungen gegen mich anhängig sind,
- weder früher noch gleichzeitig ein Habilitationsverfahren durchgeführt oder angemeldet wurde,
- die vorgelegte Habilitationsschrift ohne fremde Hilfe verfasst, die beschriebenen Ergebnisse selbst gewonnen wurden, sowie die verwendeten Hilfsmittel, die Zusammenarbeit mit anderen Wissenschaftlerinnen oder Wissenschaftlern und technischen Hilfskräften und die Literatur vollständig angegeben sind,
- mir die geltende Habilitationsordnung bekannt ist

.....
Datum

.....
Unterschrift

Inhalt

1	Einleitung und Fragestellungen	1
2	Systematische Übersichtsarbeiten randomisierter Studien - Anwendungsbeispiele	5
2.1	Hintergrund	5
2.1.1	Allgemeine Anmerkungen zu den Anwendungsbeispielen	5
2.1.2	Migräneprophylaxe mit Propranolol, Metoprolol und Flunarizin	6
2.1.3	Behandlung von Depressionen mit Hypericumextrakten	7
2.2	Methodik	8
2.2.1	Migräneprophylaxe mit Propranolol, Metoprolol und Flunarizin	8
2.2.1.1	Ein- und Ausschlußkriterien	8
2.2.1.2	Literatursuche	8
2.2.1.3	Selektionsprozess	9
2.2.1.4	Extraktion und Beurteilung	9
2.2.1.5	Zusammenfassung der Primärstudienenergebnisse	13
2.2.2	Behandlung von Depressionen mit Hypericumextrakten	15
2.2.2.1	Ein- und Ausschlusskriterien	15
2.2.2.2	Literatursuche	16
2.2.2.3	Selektionsprozess	16
2.2.2.4	Extraktion und Beurteilung	16
2.2.2.5	Zusammenfassung der Primärstudienenergebnisse	17
2.3	Ergebnisse	18
2.3.1	Migräneprophylaxe mit Propranolol, Metoprolol und Flunarizin	18
2.3.1.1	Literatursuche und Selektion	18
2.3.1.2	Allgemeine Studienmerkmale	19
2.3.1.3	Studienqualität	20
2.3.1.4	Studienergebnisse	22
2.3.2	Behandlung von Depressionen mit Hypericumextrakten	31
2.3.2.1	Literatursuche und Selektion	31

2.3.2.2	Studienmerkmale	31
2.3.2.3	Studienqualität	31
2.3.2.4	Studienergebnisse	32
2.4	Diskussion	38
2.4.1	Migräneprophylaxe mit Propranolol, Metoprolol und Flunarizin	38
2.4.2	Behandlung von Depressionen mit Hypericumextrakten	41
3	Empirisch-methodische Untersuchungen zur Studienqualität und deren Einfluss auf das Studienergebnis	43
3.1	Hintergrund	43
3.2	Methodik	45
3.2.1	Methodische Qualität randomisierter Studien in den Bereichen Akupunktur, Phytotherapie und Homöopathie	45
3.2.1.1	Studienselektion	45
3.2.1.2	Beurteilung der Studienqualität	46
3.2.1.3	Auswertung	47
3.2.2	Einfluss von Aspekten der methodischen Qualität auf das Studienergebnis am Beispiel der placebokontrollierten Studien zur Homöopathie	48
3.2.2.1	Studienselektion	48
3.2.2.2	Beurteilung der methodischen Qualität	48
3.2.2.3	Auswertung	49
3.2.3	Einfluss von Qualitätsmerkmalen auf Therapieerfolgswerten in Studien unterschiedlicher Designs am Beispiel Akupunktur bei chronischen Kopfschmerzen	50
3.2.3.1	Studienselektion	50
3.2.3.2	Literatursuche	50
3.2.3.3	Extraktion und Beurteilung	51
3.2.3.4	Auswertung	52

3.3	Ergebnisse	53
3.3.1	Methodische Qualität randomisierter Studien in den Bereichen Akupunktur, Phytotherapie und Homöopathie	53
3.3.2	Einfluss von Aspekten der methodischen Qualität auf das Studienergebnis am Beispiel der placebokontrollierten Studien zur Homöopathie	57
3.3.3	Einfluss von Qualitätsmerkmalen auf Therapieerfolgswerten in Studien unterschiedlicher Designs am Beispiel Akupunktur bei chronischen Kopfschmerzen	60
3.4	Diskussion	66
3.4.1	Methodische Qualität randomisierter Studien in den Bereichen Akupunktur, Phytotherapie und Homöopathie	66
3.4.2	Einfluss von Aspekten der methodischen Qualität auf das Studienergebnis am Beispiel der placebokontrollierten Studien zur Homöopathie	68
3.4.3	Einfluss von Qualitätsmerkmalen auf Therapieerfolgswerten in Studien unterschiedlicher Designs am Beispiel Akupunktur bei chronischen Kopfschmerzen	70
4	Analyse der vorliegenden systematischen Übersichtsarbeiten zu Akupunktur, Phytotherapie und Homöopathie	73
4.1	Hintergrund	73
4.2	Methodik	74
4.2.1	Ein- und Ausschlusskriterien	74
4.2.2	Literatursuche	75
4.2.3	Selektionsprozess	75
4.2.4	Extraktion und Beurteilung	76
4.2.5	Vergleich der Übersichtsarbeiten zu gleichen Fragestellungen	77
4.3	Ergebnisse	78
4.3.1	Zusammenstellung der systematischen Übersichtsarbeiten zur Akupunktur	78
4.3.2	Zusammenstellung der systematischen Übersichtsarbeiten zur Phytotherapie	84
4.3.3	Zusammenstellung der systematischen Übersichtsarbeiten zur Homöopathie	93
4.3.4	Merkmale und Qualitätsbewertung der Übersichtsarbeiten zu allen drei Therapien	97

4.3.5	Vergleich der Übersichtsarbeiten zu gleichen Fragestellungen	102
4.4	Diskussion	109
4.4.1	Zusammenstellung der systematischen Übersichtsarbeiten zu Akupunktur, Phytotherapie und Homöopathie	109
4.4.2	Merkmale und Qualitätsbewertung der Übersichtsarbeiten zu allen drei Therapien	112
4.4.3	Vergleich der Übersichtsarbeiten zu gleichen Fragestellungen	114
5	Schlussbetrachtung	116
6	Tabellenanhang	121
6.1	Tabellen zu den Studien zu Propranolol, Metoprolol und Flunarizin zur Migräneprophylaxe	121
6.2	Tabellen zu den Studien zu Hypericumextrakten bei Depression	147
7	Erläuterungen zum Verständnis der Effektmaßdarstellungen	155
8	Literaturverzeichnis	159
9	Zusammenfassung	201
	Nennung der Wissenschaftler, die bei einzelnen systematischen Übersichtsarbeiten bzw. Untersuchungen mitgewirkt haben	205
	Danksagung	206
	Publikationen zu in der vorliegenden Habilitationsschrift präsentierten Daten	207

1 Einleitung und Fragestellungen

Angesichts der Flut von Forschungsergebnissen, die jedes Jahr publiziert werden, ist es selbst in abgegrenzten medizinischen Fachgebieten für den Einzelnen kaum noch möglich, die Übersicht zu behalten. Für den Arzt, der bei der Behandlung seiner Patienten eigentlich die beste externe Evidenz in seine Entscheidungsfindung miteinbeziehen sollte, ist es unmöglich, über ein weites Krankheitsspektrum hinweg alle relevanten klinischen Studien zu lesen und kritisch zu interpretieren. In diesem Zusammenhang haben in den letzten Jahren systematische Übersichtsarbeiten große Bedeutung erlangt. Übersichtsarbeiten (im folgenden synonym mit der Bezeichnung Review gebraucht) werden dann als systematisch bezeichnet, wenn sie vordefinierte und transparente Methoden bzgl. Literatursuche, Studienselektion und Studienbewertung verwenden (73). Systematische Übersichtsarbeiten, bei denen die einbezogenen Studien (Primärstudien) in einer integrierenden statistischen Analyse wie eine einzige große Studie ausgewertet (gepoolt) werden, bezeichnet man als Meta-Analysen. Als narrativ werden demgegenüber traditionelle Übersichtsarbeiten bezeichnet, die nicht nach systematischen und explizit beschriebenen Methoden vorgehen.

Obwohl systematische Übersichtsarbeiten grundsätzlich in allen Bereichen der Forschung durchführbar sind, findet man sie besonders häufig zu randomisierten Interventionsstudien. Die Fragestellungen systematischer Reviews gleichen im Prinzip denjenigen der Primärstudien. Sie sollten explizit in Bezug auf die Indikation bzw. die Patienten, auf Intervention, Kontrollintervention und nach Möglichkeit auf die Zielparameter formuliert sein (z.B. „Senkt Flunarizin in der prophylaktischen Behandlung von Patienten mit Migräne die Anzahl der Migräneattacken in ähnlichem Ausmaß wie Propranolol?“). Allerdings ist eine genaue Definition der Zielparameter in der Fragestellung bei Erkrankungen, die nicht ohne weiteres in einer einzigen Größe zusammengefaßt werden können, häufig schwierig oder wenig sinnvoll, sodass die Fragestellung in der Realität diesbezüglich häufig offengehalten werden muss (z.B. „Ist Flunarizin bei der prophylaktischen Behandlung von Migränepatienten ähnlich wirksam wie Propranolol?“).

Aus einer konkreten Fragestellung ergibt sich das erste entscheidende methodische Element einer systematischen Übersichtsarbeit: die Ein- und Ausschlusskriterien. Diese sollten grundsätzlich Angaben zu Patienten bzw. Erkrankung (z.B. „Migräne mit und ohne Aura“), Prüfintervention (z.B. „Flunarizin“), Kontrollintervention (z.B. „Placebo oder Betablocker“), Design (z.B. „ausschließlich randomisierte, doppelblinde Studien“) und Zielparametern machen (z.B. „mindestens einer der folgenden Zielparameter wurde gemessen: Attackenfrequenz, Kopfschmerztag, Anzahl Responder“). Darüberhinaus kann es sinnvoll sein, weitere Ein- und Ausschlusskriterien zu formulieren (z.B. „ausgeschlossen werden Studien mit einer Beobachtungsdauer von weniger als 8 Wochen“).

Die Literatursuche muß im nächsten Schritt gewährleisten, dass alle relevanten Studien, die möglicherweise die Einschlusskriterien erfüllen (= Primärstudien) auch tatsächlich identifiziert werden. Grundsätzlich herrscht in der Fachdiskussion Konsens, dass die Literatursuche so umfassend wie möglich erfolgen sollte (271), d.h. dass neben gängigen Datenbanken wie *Medline* oder *Embase* auch andere Quellen herangezogen und dass grundsätzlich Studien in allen Sprachen berücksichtigt werden sollten (das Letztere gehört eigentlich zu den Selektionskriterien, ist praktisch aber vor allem für Literatursuche und Extraktion relevant). Umfassende Suchen sind jedoch aufwändig und es ist nicht klar, ob der zusätzliche Aufwand tatsächlich die erwünschte Mehrinformation bzw. bessere Information bringt (332). So wurde gezeigt, dass nicht über Medline identifizierbare Studien und insbesondere unpublizierte Studien häufig schlechte Qualität haben (99;101;282). Ohne eine umfassende Suche bleibt im Einzelfall jedoch eine erhebliche Unsicherheit, ob die gefundenen Studien tatsächlich repräsentativ sind. Eine umfassende Literatursuche kann allerdings nicht alle Probleme – v.a. den sogenannten „publication bias“, d.h. die Nichtpublikation unerwünschter Ergebnisse – mit Sicherheit ausschließen. Sie erlaubt jedoch „Sensitivitätsanalysen“, d.h. Analysen, in denen empirisch untersucht werden kann, ob z.B. die Ergebnisse in Zeitschriften publizierter Arbeiten positiver sind, als die von an anderer Stelle oder nicht publizierten.

Ein häufig wenig beachteter und in Veröffentlichungen von systematischen Übersichtsarbeiten wenig beschriebener Schritt ist die Umsetzung der Studienselektion. Hier sollte für den Leser gut nachvollziehbar sein, wie aus den zunächst durch die Literatursuche identifizierten Arbeiten die letztlich eingeschlossenen ausgewählt wurden. Eine Liste der Studien, die den Einschluss „knapp verfehlt“ haben, mit Ausschlussgründen ist hier unabdingbar. Unterschiede in den Ergebnissen systematischer Übersichtsarbeiten ergeben sich nämlich häufig aus einer Mixtur subtiler Unterschiede in den Einschlusskriterien und Literatursuchen (202) (siehe Kapitel 4).

Die Extraktion von Information aus den Primärstudien und deren Bewertung sollte nach Möglichkeit von mindestens zwei Beurteilern (Reviewer) durchgeführt werden (102). Für die Informationsextraktion bieten sich Formulare oder vordefinierte Tabellenmatrices an, in die die jeweils relevanten Informationen standardisiert eingefügt werden. Die Beurteilung der Qualität der Studien sollte nach vordefinierten, anerkannten Kriterien erfolgen. Dieser relevante, aber problematische Schritt wird in der vorliegenden Monographie in Kapitel 3 detailliert diskutiert.

Schließlich mündet der Reviewprozeß in den letztlich entscheidenden Schritt: Die Zusammenfassung der Ergebnisse. Sind die Primärstudien vergleichbar und präsentieren diese ihre Ergebnisse ausreichend detailliert, ist die Berechnung von Effektgrößen – d.h. von Maßen, die für die einzelne Studie die Größe des Unterschiedes zwischen den verglichenen Gruppen quantifizieren – eindeutig die Methode der Wahl (102;364). Hier ist nochmals darauf hinzuweisen, dass auch bei dieser Stufe noch nicht von einer Meta-Analyse zu sprechen ist. Erst wenn die

Effektgrößen der einzelnen Studien „gepoolt“ werden, d.h. integrierend statistisch wie Daten aus einer einzigen großen Studie ausgewertet werden, handelt es sich um eine Meta-Analyse. Der Großteil der Kritik an systematischen Übersichtsarbeiten (siehe z.B. (130;132)) bezieht sich auf diesen letzten Schritt, dessen Durchführung sorgfältig zu erwägen ist und dessen Ergebnisse meist nur mit großer Vorsicht zu interpretieren sind. In vielen Fällen scheitert jedoch bereits die Berechnung von Effektgrößen für die einzelnen Studien an der enormen Heterogenität oder der insuffizienten Berichterstattung der Primärstudien. In diesen Fällen ist dann lediglich eine deskriptive Zusammenfassung der Primärstudienenergebnisse möglich. Häufig wird auch eine Kategorisierung der Ergebnisse (z.B. in „positiv“ und „negativ“, ein sogenannter „vote count“) durchgeführt, der jedoch äusserst unpräzise und irreführend sein kann. Dem Leser sollte klar sein, dass das Zählen „positiver“ und „negativer“ Ergebnisse letztlich nichts anderes als eine primitive Form der Meta-Analyse darstellt, auch wenn derartige systematische Übersichtsarbeiten gelegentlich als „qualitativ“ bezeichnet werden. Gute Reviews heterogener Studien enthalten daher meist ausführliche Tabellen, bei denen es dem kritischen Leser nicht erspart bleibt, die deskriptiv zusammengefassten Primärstudienenergebnisse durchzusehen.

Angesichts der hier angedeuteten Probleme ist es wenig überraschend, dass systematische Übersichtsarbeiten trotz ihrer großen Bedeutung häufig Gegenstand von kontroversen Debatten sind (130;132). Darüberhinaus kann die methodische Entwicklung der vergleichsweise jungen Technik keineswegs als abgeschlossen bezeichnet werden (100). Zu den einzelnen Schritten der Reviewmethodik werden in den letzten Jahren zunehmend empirische Untersuchungen durchgeführt, um die Auswirkungen einzelner Vorgehensweisen bzw. Einflüsse verschiedener Störfaktoren zu untersuchen. Eine spezielle Datenbank, in der diese Untersuchungen registriert werden, ist bereits verfügbar (Cochrane Methodology Register in der Cochrane Library, Oxford: Update Software).

Die vorliegende Monographie präsentiert 1. zwei größere eigene systematische Übersichtsarbeiten, 2. empirisch-methodische Untersuchungen zum Thema Studienqualität und deren Auswirkung auf das Ergebnis und 3. eine Analyse der vorliegenden systematischen Übersichtsarbeiten in den Bereichen Akupunktur, Phytotherapie und Homöopathie. Sie versteht sich als Beitrag zur methodischen Diskussion und Entwicklung von systematischen Übersichtsarbeiten.

Folgende Fragestellungen werden bearbeitet:

1. Sind Propranolol, Metoprolol und Flunarizin bei der Migräneprophylaxe (Vorbeugung von Migräneattacken) wirksamer als Placebo und ähnlich wirksam oder wirksamer als andere Medikamente? (Kapitel 2)
2. Sind Johanniskrautextrakte (*Hypericum perforatum* L.) bei der Behandlung von Depressionen wirksamer als Placebo und ähnlich wirksam wie synthetische Antidepressiva sowie ggfs. nebenwirkungsärmer? (Kapitel 2)
3. Wie ist die methodische Qualität randomisierter Studien zu den Verfahren Akupunktur, Phytotherapie und Homöopathie? (Kapitel 3)
4. Wie wirken sich Unterschiede in der Qualität von Studien auf die Ergebnisse aus? (Kapitel 3)
5. Welche systematischen Übersichtsarbeiten liegen in den Bereichen Akupunktur, Phytotherapie und Homöopathie vor? (Kapitel 4)
6. Kommen unterschiedliche systematische Übersichtsarbeiten zu gleichen Fragestellungen zu denselben Ergebnissen? (Kapitel 4)

Anmerkung: Kapitel 2 präsentiert zahlreiche Abbildungen zur Zusammenfassung von Studienergebnissen, die Bezeichnungen in englischer Sprache enthalten. Diese Abbildungen wurden mit der Cochrane Collaboration Review Manager Software, Version 4.1. erstellt und sind nicht editierbar, sodass die Bezeichnungen nicht ins Deutsche übersetzt werden konnten.

2 Systematische Übersichtsarbeiten randomisierter Studien - Anwendungsbeispiele

2.1 Hintergrund

2.1.1 Allgemeine Anmerkungen zu den Anwendungsbeispielen

Die für dieses Kapitel gewählten Anwendungsbeispiele zu randomisierten Studien zur Migräneprophylaxe mit Propranolol, Metoprolol und Flunarizin sowie zur Behandlung von Depressionen bzw. von depressiven Verstimmungen mit Hypericumextrakten sind aufgrund der Anzahl der einbezogenen Primärstudien und deren Heterogenität im Vergleich zu vielen publizierten Übersichtsarbeiten sehr umfangreich und komplex.

Die große Zahl von Primärstudien ist aus methodischer Sicht von Vorteil, da sie aussagekräftige Untersuchungen zur Qualität und Sensitivitätsanalysen im Rahmen der quantitativen Auswertungen überhaupt erst ermöglicht.

Die Komplexität ergibt sich einerseits aus den bearbeiteten Krankheitsbildern: Weder Migräne noch Depression sind klinisch in ihrer Ausprägung und im Verlauf leicht zu beurteilen. In Studien erfolgt dies bei der Migräne am besten über Tagebücher, in denen eine ganze Reihe von Parametern (z.B. Attackenhäufigkeit, -schwere, Medikationsverbrauch, Begleitsymptome) erfaßt wird (198). Die Präsentation der Ergebnisse wird von verschiedenen Autoren unterschiedlich gehandhabt (z.B. Mittelwert oder Median, Standardabweichung, Standardfehler oder 95%-Konfidenzintervall, Attackenrückgang im Vergleich zur Baselinephase in Prozent oder absolut oder Attackenhäufigkeit in den einzelnen Phasen etc.) und ist außerdem sehr häufig unvollständig. Bei der Depression werden typischerweise Fremd- und Selbstbeurteilungsinstrumente eingesetzt, in denen eine Reihe von Symptomen bewertet und quantifiziert werden. Zwar werden bei vielen Instrumenten auch „Responder“ und „Nonresponder“ kategorisiert, d.h. Patienten die unter Therapie eine bestimmte Verbesserung erfahren oder nicht, jedoch sind solche Kategorisierungen mit einem Informationsverlust verbunden.

Ein weiterer Grund für die Komplexität der folgenden systematischen Übersichtsarbeiten sind die Interventionen selbst. Propranolol, Metoprolol und Flunarizin sind „alte“ Substanzen; dementsprechend stammen auch die meisten Studien aus einer Zeit, in denen wesentlich geringere Anforderung an die Qualität gestellt wurden und heute übliche allgemeine (Good Clinical Practice Richtlinien für die Durchführung klinischer Studien) und migränespezifische (197;198) Standardempfehlungen noch nicht verfügbar waren. Berichterstattung und Zuverlässigkeit der Daten sind daher häufig suboptimal. Hypericumextrakte wurden vor allem in den neunziger Jahren Untersuchungsgegenstand randomisierter Studien. Durchführungs- und Berichtsqualität sind daher

in der Regel besser. Die unklare Vergleichbarkeit unterschiedlicher Hypericumextrakte – ein typisches Problem bei Übersichtsarbeiten zu Phytotherapeutika – stellt hier die Durchführbarkeit einer quantitativen Meta-Analyse in Frage. Ein weitere Schwierigkeit ist, dass Hypericumextrakte primär von niedergelassenen Ärzten bei Patienten mit leichten bis mittelschweren Depressionen angewendet werden, einer Klientel, die – obwohl zahlenmäßig sehr groß – in der klinischen Forschung zu konventionellen Antidepressiva kaum berücksichtigt wird (232). Dadurch ist der Vergleich mit diesen Substanzen erschwert.

Die Darstellung der Anwendungsbeispiele geht explizit auf diese von vornherein bei der Planung erwarteten Schwierigkeiten ein.

2.1.2 Migräneprophylaxe mit Propranolol, Metoprolol und Flunarizin

Migräne ist eine weit verbreitete Erkrankung mit einer geschätzten Prävalenz zwischen 8 und 16% in westlichen Industriestaaten (150;383). Bei der Mehrzahl der Patienten ist es ausreichend, ausschließlich die Migräneattacken zu behandeln (87). Bei einer nicht zu vernachlässigenden Minderheit der Patienten ist jedoch auch eine prophylaktische bzw. Intervallbehandlung indiziert, da die Attacken zu häufig oder zu intensiv sind, oder schlecht auf die Akutbehandlung ansprechen. Als Medikamente erster Wahl sind in Leitlinien (21;87) die Betablocker Propranolol und Metoprolol und der Calciumantagonist Flunarizin angegeben.

Eine große Zahl randomisierter Studien wurde zu diesen Wirkstoffen durchgeführt. Aktuelle systematische Übersichtsarbeiten liegen jedoch nicht in einfach zugänglicher Form vor. Zu Propranolol wurde von Holroyd et al. 1991 eine Meta-Analyse publiziert (189). Die Autoren schlossen, dass die kurzfristige Wirksamkeit über Placebo hinaus gut belegt sei, jedoch nur wenig Information zu Langzeiteffekten verfügbar sei. Diese Übersicht wird heutigen Anforderungen jedoch nicht gerecht, vor allem weil in der Meta-Analyse nicht die Zwischengruppenunterschiede (z.B. Differenz Verum – Placebo) sondern nur die Innergruppenunterschiede (vorher – nachher) verwendet wurden und weil die Bewertung ausschließlich auf der Basis der heute in der Migränerforschung nicht mehr empfohlenen (198) Kopfschmerzindices (Scoremaße, in die meist Häufigkeit und Intensität sowie evtl. andere Größen mit unterschiedlicher Gewichtung eingingen) erfolgte. Im Rahmen einer umfangreichen Bewertung aller verfügbaren Optionen zur Migräneprophylaxe bewertete eine Gruppe der Duke University in Durham (USA) im Auftrag der Agency of Health Care Policy and Research (AHCPR) auch Propranolol, Metoprolol und Flunarizin und klassifizierte die Wirksamkeit dieser Substanzen als belegt (62). Diese umfassende Zusammenstellung entspricht zwar den gängigen Anforderungen an systematische Übersichtsarbeiten, ist jedoch nur in Berichtsform publiziert und in Bezug auf die einzelnen bearbeiteten Interventionen recht unübersichtlich. Aus diesem Grund (sowie wegen des für

methodische Fragestellungen interessanten Potenzials dieses Studiensets) wurde eine systematische Übersichtsarbeit entsprechend gängiger Vorgehensweisen (2) zu den randomisierten Studien zu Propranolol, Metoprolol und Flunarizin durchgeführt. Folgende Fragen sollten beantwortet werden:

Ist durch randomisierte Studien belegt, dass

1. Propranolol, Metoprolol und Flunarizin zur Prophylaxe von Migräneattacken bzw. zur Intervallbehandlung der Migräne wirksamer sind als Placebo?
2. Propranolol, Metoprolol und Flunarizin zur Prophylaxe von Migräneattacken bzw. zur Intervallbehandlung der Migräne gleich wirksam oder wirksamer sind als andere Medikamente?

Außerdem sollte untersucht werden, wie die Qualität der Studien (aus der Sicht der Methodik klinischer Studien allgemein und in Bezug auf die Interpretierbarkeit für die Routinetherapie der Migräne) einzuschätzen ist und wie häufig Nebenwirkungen im Vergleich zu Placebo und anderen Medikamenten sind.

2.1.3 Behandlung von Depressionen mit Hypericumextrakten

Extrakte aus *Hypericum perforatum* (Johanniskraut) sind in Deutschland die am häufigsten verschriebenen Medikamente zur Behandlung von Depressionen bzw. von depressiven Verstimmungen (483). Johanniskrautextrakte enthalten mindestens 10 pharmakologisch potentiell relevante Wirkstoffgruppen, darunter Naphthodianthrone (z.B. Hypericin, auf dessen Gehalt zahlreiche marktgängige Präparate standardisiert sind), Flavonoide (z.B. Quercetin), Xanthone und Biflavonoide (340). Der exakte Wirkungsmechanismus der häufig beschriebenen antidepressiven Effekte ist jedoch nicht vollkommen klar (318).

Seit den achtziger Jahren wurde eine ganze Reihe randomisierter Studien zur Behandlung von Depressionen mit Johanniskrautextrakten durchgeführt. Unter der Leitung des Autors hat eine Arbeitsgruppe 1996 eine erste Meta-Analyse dieser Arbeiten vorgelegt (286); die Metaanalyse wurde 1988 aktualisiert (285). Da in den vergangenen Jahren weitere und vor allem methodisch bessere randomisierte Studien durchgeführt und publiziert wurden, wurde eine erneute Aktualisierung durchgeführt.

Folgende Fragestellungen sollten beantwortet werden.

1. Sind Hypericumextrakte bei Behandlung von Depressionen bzw. von depressiven Verstimmungen wirksamer als Placebo?
2. Sind Hypericumextrakte ähnlich effektiv wie synthetische Antidepressiva?
3. Sind Hypericumextrakte nebenwirkungsärmer als synthetische Antidepressiva?

2.2 Methodik

2.2.1 Migräneprophylaxe mit Propranolol, Metoprolol und Flunarizin

2.2.1.1 Ein- und Ausschlusskriterien

Um in die systematische Übersichtsarbeit eingeschlossen zu werden, musste eine Studie folgende Kriterien erfüllen:

1. Patienten: Patienten (im Durchschnitt über 18 Jahre) mit Migräne (mit oder ohne andere Kopfschmerzen). Studien mit Migräne- und anderen Kopfschmerzpatienten wurden nur aufgenommen, wenn separate Ergebnisse zu den Migränepatienten vorlagen.
2. Interventionen: In der Prüfgruppe Propranolol, Metoprolol oder Flunarizin zur Intervallbehandlung, in der Kontrollgruppe Placebo oder eine andere medikamentöse Migräneprophylaxe. Studien zu verschiedenen Dosen oder Applikationsformen der gleichen Substanz wurden nur einbezogen, wenn zusätzlich auch ein Vergleich mit einer der zwei oben genannten Arten von Kontrollen erfolgte. Studien zur Behandlung akuter Migräneattacken wurden ausgeschlossen.
3. Design: explizit randomisierte und quasirandomisierte (z.B. Zuteilung entsprechend Geburtsdatum) klinische Studien sowie doppelblinde Studien, in denen neben den Patienten die aufnehmende und behandelnde Person verblindet war (dieses Kriterium wurde einbezogen, da in älteren Doppelblindstudien die Randomisation häufig nicht explizit erwähnt ist, die Doppelblindbedingungen jedoch nur mit Hilfe einer zumindest quasirandomisierten Zuteilung möglich sind).
4. Zielkriterien: Beurteilung mindestens eines der folgenden Parameter nach mindestens 4 Wochen Beobachtungsdauer: Globalbeurteilung der Wirksamkeit (Response), Kopfschmerzindex, Zahl der Kopfschmerztage, Anfallshäufigkeit, -intensität. Studien, die ausschließlich andere Parameter beurteilten oder nach Therapiebeginn weniger als 4 Wochen beobachteten, wurden ausgeschlossen.

2.2.1.2 Literatursuche

Zur möglichst umfassenden Identifikation der verfügbaren Studien wurden folgende Quellen herangezogen:

1. Medline (WinSpis 4.0) 1966-2000/4; Vorgehensweise: Und-Verknüpfung des Suchbegriffs für die Erkrankung (migraine) mit den Begriffen für die Therapie (flunarizin*, propranolol or propranolol, metoprolol) sowie für kontrollierte klinische Studien (Strategie aus Cochrane Handbuch (2) Stufen 1-25)
2. Cochrane Controlled Trial Register (2000, issue 2): Migraine and Therapiebegriff (siehe Medlinesuche)

3. Prüfen der gefundenen Studien auf relevante Querverweise
4. Prüfen von Übersichtsarbeiten auf relevante Querverweise
5. Regelmäßige Handsuchen in den Zeitschriften Cephalalgia, Headache und Pain seit Januar 1999 sowie Screeningsuchen über PubMed (migraine and Therapiebegriff) im Frühjahr 2001 zur Identifikation neuer Studien

2.2.1.3 Selektionsprozess

Die primären Suchergebnisse aus Medline und dem Cochrane Controlled Trial Register wurden dokumentiert und am Bildschirm einzeln geprüft (sofern ein Abstract vorhanden war, erfolgte immer eine Durchsicht des Abstracts). In einem ersten Screeningschritt wurden alle Referenzen aussortiert, bei denen eindeutig erkennbar war, dass sie die Einschlusskriterien nicht erfüllten (z.B. Studien an Kindern, placebokontrollierte Studien zu Sumatriptan, bei denen lediglich das Wort Propranolol im Abstract vorkam etc.). Alle übrigen Referenzen wurden in eine Datenbank (Reference Manager) eingelesen, im Volltext besorgt und formal (möglichst von zwei Beurteilern) mit einem Standardbogen auf die Erfüllung der Einschlusskriterien geprüft. Die Ausschlussgründe wurden jeweils dokumentiert. Die bibliographischen Angaben der geprüften Übersichtsarbeiten sowie die Querverweise auf potentiell relevante Studien in den besorgten Publikationen wurden ebenfalls geprüft. Lag die entsprechende Publikation nicht vor, wurde sie im Volltext besorgt.

2.2.1.4 Extraktion und Beurteilung

Die Extraktion von bibliographischen Angaben, Information zu Patienten, Methodik, Ergebnissen und Schlußfolgerungen sowie die Beurteilung der Studienqualität erfolgte mit einem vorgeprüften Formular durch einen oder zwei Reviewer. Ergebnisdaten wurden nach Möglichkeit zu vier Zeitpunkten (vor Behandlung, bis zu 4 Wochen nach Behandlungsende, 5 Wochen bis 3 Monate nach Behandlungsende (early follow-up) und mehr als 3 Monate nach Behandlungsende (late follow-up) zu folgenden Parametern extrahiert: Response, Zahl der Kopfschmerztag, Attackenhäufigkeit, -intensität, Kopfschmerzindex, sowie der Verbrauch von Akutmedikation. Wenn keine verwertbaren Daten (d.h. Daten, mit deren Hilfe ein Effektstärkemaß berechnet werden konnte, siehe 2.2.1.5.) vorlagen, wurde dokumentiert, ob entweder der Parameter zwar gemessen, die Ergebnisse aber nicht ausreichend oder gar nicht präsentiert wurden, der Parameter nicht gemessen wurde, oder ob die Situation jeweils unklar war.

Die Bewertung der *methodischen Qualität* erfolgte mit Hilfe des Jadadscores (206) und der Delphi Liste (469). Ziel beider Instrumente ist die Beurteilung von Aspekten bzgl. des Designs und der Durchführung von klinischen Studien, die relevant sind für ein verzerrungsfreies Studienergebnis.

Der Jadadscore ist eine systematisch entwickelte und validierte Skala zur Beurteilung der methodischen Qualität bzw. der „quality of reporting“ bezüglich methodischer Schlüsselkriterien. Er hat drei Items, bei denen maximal fünf Punkte (2-2-1) vergeben werden.

1. Randomisation:

Wenn die Studie explizit als randomisiert bezeichnet ist, wird ein Punkt vergeben. Ein Zusatzpunkt wird vergeben, wenn zusätzlich die Methode zur Generierung der Randomisationssequenz beschrieben ist und diese Methode adäquat ist (z.B. Computer-generierte Randomliste, Random-Table etc.). Erfolgte die Zuteilung nach Geburtsdatum, Wochentag u.ä., wird kein Punkt vergeben (auch wenn die Studie angeblich „randomisiert“ ist).

2. Doppelblindbedingungen:

Wenn die Studie als doppelblind (Patienten und Ärzte verblindet) beschrieben ist, wird ein Punkt vergeben. Wenn die Doppelblindbedingungen genauer und glaubhaft beschrieben sind (z.B. mit einem Statement, daß Verum und Placebo „identical in taste and appearance“ waren), wird ein zusätzlicher Punkt gegeben. Wenn dagegen aus der Beschreibung klar wird, daß Prüf- und Vergleichsbehandlung unterscheidbar waren, gibt es keinen Punkt, auch wenn die Autoren behaupten, die Studie sei doppelblind gewesen.

3. Drop-outs/Abbrüche: Hier wird ein Punkt vergeben

- wenn explizit und glaubhaft gesagt wird, daß alle Patienten, die randomisiert wurden, auch tatsächlich analysiert wurden
- oder wenn für *alle Gruppen Zahl und Gründe* für Abbrüche und Ausschlüsse genannt sind (Wenn z.B. nur die Zahl genannt wird oder Zahl und Gründe zwar genannt werden, aber nicht für die Gruppen separat, sondern nur insgesamt, wird kein Punkt vergeben)

Bei den Originalinstruktionen ist mit Doppelblindbedingungen die Verblindung von Patienten und *Beurteilern* gemeint (d.h., es können auch nicht-medikamentöse Studien unter bestimmten Bedingungen doppelblind sein). Da im vorliegenden Review medikamentöse Interventionen bearbeitet wurden, wurde die klassische Definition der Doppelblindbedingungen (Patient, Therapeut und Beurteiler verblindet) zugrundegelegt.

Die 9 Fragen umfassende Delphi-Liste wurde von einem internationalen Team von Methodikern in einem Delphiprozess (d.h. mehreren Befragungs- und Konsensusrunden) entwickelt (469). Die Fragen können jeweils mit „ja“, „nein“ oder „unklar“ beantwortet werden. Da es für die Anwendung der Liste keine weitergehenden Instruktionen von den Autoren gibt (persönliche Mitteilung A. Verhagen), wurden eigene Vorgaben für die konkrete Anwendung einiger Kriterien formuliert. Die Fragen lauten:

1. Erfolgte die Gruppenzuteilung randomisiert?

2. Erfolgte die Gruppenzuteilung verblindet (concealed)?

Weitergehende Instruktion: Geben Sie Antwort „ja“, wenn beschrieben ist, wie gewährleistet wurde, daß der aufnehmende Arzt nicht wußte, welche Behandlung der nächste Patient bekommen würde. Als adequat ist z.B. anzusehen, wenn „serially numbered, identical medication containers“ (der Patient bekommt in einer doppelblinden Studie jeweils die nächste Nummer zugeteilt, die je nach Randomisation Verum oder Placebo etc. entspricht), eine „central telephone randomisation“ oder „sealed opaque envelopes“ (nur relevant bei nichtmedikamentösen Verfahren) benutzt wurden.

3. Waren die Gruppen vor der Behandlung bzgl. der relevantesten prognostischen Indikatoren vergleichbar?

Weitergehende Instruktion: Geben Sie Antwort „ja“, wenn mindestens zu Alter, Geschlecht, Erkrankungsdauer, Symptomintensität und ggf. Diagnosen (bei Studien, bei denen die Patienten zusätzlich andere Kopfschmerzen haben) a) Daten präsentiert wurden und b) diese in beiden Gruppen ausreichend ähnlich waren. Angaben ausschließlich zur Gesamtgruppe in cross-over Studien wurden als nicht ausreichend betrachtet.

4. Wurden Ein- und Ausschlusskriterien beschrieben?

5. Erfolgte die Verlaufsbeurteilung (outcome assessment) unter Blindbedingungen?

Weitergehende Instruktion: Für eine Antwort „ja“ sollte zusätzlich zum Patienten, der bei Kopfschmerzstudien im allgemeinen selbst der wichtigste Beurteiler ist, auch die begleitende Person (in der Regel der Arzt), die die Tagebücher einsammelt oder eine Beurteilung vornimmt, verblindet sein.

6. War der Therapeut verblindet?

7. War der Patient verblindet?

Weitergehende Instruktion: Wird eine Medikamentenstudie als doppelblind bezeichnet, sollten die Frage 5-7 mit „ja“ beantwortet werden (es sei denn, es gibt klare Hinweise, dass eine oder mehrere Gruppen doch nicht verblindet waren).

8. Wurden die Ergebnisse für die Hauptzielparameter ausreichend detailliert beschrieben?

Weitergehende Instruktion: Antwort „ja“ ankreuzen wenn für das oder die wichtigsten outcome-Maße (typischerweise Kopfschmerztage, -häufigkeit, -intensität, -indices) ein zentraler Schätzer (Mittelwert oder Median) *und* ein Maß zur Variabilität angegeben ist (Standardabweichung, Quartile etc.) bzw. ein Maß, aus dem die Variabilität rückbestimmt werden kann (Standardfehler, Konfidenzintervall, exakter statistischer Testwert).

9. Wurde eine Intent-to-treat-Analyse durchgeführt?

Weitergehende Instruktion: Antwort „ja“ ankreuzen, wenn entweder explizit mehr als 90% der randomisierten Patienten auch tatsächlich analysiert wurden oder eine intent-to-treat Analyse durchgeführt wurde.

Zusätzlich zur Bewertung der methodischen Qualität wurde auch die *Angemessenheit von Beobachtung und Berichterstattung* mithilfe einer selbstentwickelten Liste beurteilt. Ziel war eine Abschätzung von Aspekten der klinischen Interpretierbarkeit, d.h. es geht um die Frage, ob ein erfahrener Kopfschmerztherapeut die Studie, so wie sie präsentiert ist, sinnvoll interpretieren kann (also prüfen, ob die Ergebnisse auf seinen Patienten übertragbar sind und ggf. eine Prognose erlauben). Die Liste umfasst 10 Fragen, die jeweils mit „ja“ und „nein“ beantwortet werden konnten:

1. Wurde die Patientenselektion beschrieben?

Antwort „ja“ ankreuzen, wenn zumindest ansatzweise beschrieben ist, wie die Stichprobe zustande kam. Mindestanforderung sind dabei ein Statement, woher die Patienten kamen (z.B. Formulierungen wie, „die Patienten wurden von niedergelassenen Ärzten an unsere Schmerzzambulanz überwiesen“ – also Informationen zur Art des Prüfzentrums und der Rekrutierung der Patienten) sowie zumindest die Angabe basaler Ein- und Ausschlusskriterien.

2. Wurde eine klare Kopfschmerzdiagnose beschrieben?

Antwort „ja“ ankreuzen, wenn beschrieben ist, welche und wie die Kopfschmerzdiagnose gestellt wurde (also z.B. nach Kriterien der International Headache Society oder des ad-hoc-Committee oder Beschreibung bzw. Quelle für ein anderes Vorgehen).

3. Wurden die Patienten ausreichend beschrieben?

Für beide Gruppen sollten mindestens die 4 folgenden Charakteristika beschrieben sein: Kopfschmerzdiagnose (also z.B. Anzahl der Patienten mit Migräne mit oder ohne Aura), Alter (mindestens Mittelwert, Median oder Range), Geschlecht und ein Maß (z.B. Kopfschmerztage, Schmerzintensität, Kopfschmerzindex etc.), das über die Intensität der Beschwerden bei Studienbeginn Aufschluß gibt.

4. Erfolgte eine Vorbeobachtung von mindestens 4 Wochen Dauer zur Ermittlung der Beschwerdeintensität vor Behandlungsbeginn?

5. Wurden die Kointerventionen beschrieben?

Antwort „ja“ ankreuzen, wenn zumindest orientierend beschrieben ist, wieviele Patienten welche Akutmedikation in welchem Ausmaß benutzten (z.B. Akutmedikation bei 45% der Patienten Analgetika, bei 47% Triptane, bei 23% Ergotamine etc.)

6. Wurde ein Kopfschmerztagebuch oder –kalender zur Beurteilung der Beschwerdeintensität verwendet?

7. Wurden Daten zur Häufigkeit der Kopfschmerzen präsentiert?

Antwort „ja“ ankreuzen, wenn deskriptive Daten (zentrale Tendenz *und* ein Verteilungsmaß, z.B. Mittelwert und Standardabweichung) zu mindestens einem Zeitpunkt am Ende oder nach der Behandlung zur Kopfschmerzhäufigkeit (also z.B. Anzahl der Kopfschmerztage oder Attackenfrequenz) vorhanden sind. Wird ausschließlich ein Kopfschmerzindex (in dem z.B. Intensität und Kopfschmerztage multipliziert werden und dann nicht mehr separat erkennbar sind) präsentiert, bitte Antwort „nein“ ankreuzen.

8. Wurden Daten zur Schmerzintensität präsentiert?

Wie 7, jedoch in Bezug auf ein Maß zur Schmerz*intensität*.

9. Wurden 2 Monate nach Beginn der Therapie weitgehend alle Patienten beurteilt und analysiert?

Antwort „ja“ ankreuzen, wenn nach Beginn der Therapie Ergebnisse über einen Beobachtungszeitraum von mindestens 2 Monaten von mindestens 90% der randomisierten Patienten vorlagen. Wenn nicht klar ist, ob Daten von 90% der Patienten ausgewertet werden konnten, „nein“ ankreuzen.

10. Wurden 6 Monate nach Beginn der Therapie weitgehend alle Patienten beurteilt und analysiert?

Wie 9., jedoch in Bezug auf 6 Monate und mindestens 80% der randomisierten Patienten.

2.2.1.5 Zusammenfassung der Primärstudienresultate

Bei der Protokollerstellung war davon ausgegangen worden, dass aufgrund von Heterogenität und Alter der Studien, von multiplen unterschiedlich gemessenen Zielparametern und mangelhafter Datenpräsentation eine umfassende Zusammenfassung der Primärstudienresultate in Effektmaßen vermutlich nicht möglich sein würde. Diese Einschätzung bewahrheitete sich. Soweit möglich wurden jedoch Effektmaße für die Responderanzahl, die Attacken- bzw. Kopfschmerzhäufigkeit und die Anzahl der Patienten mit Nebenwirkungen berechnet. Daten zu Kopfschmerzindizes wurden nur selten ausreichend detailliert berichtet, sodass auf diese in der Folge nicht weiter eingegangen wird. Als Responderanzahl wurde, wenn vorhanden, die Anzahl der Patienten mit einer mindestens 50%igen Reduktion der Attackenzahl im Vergleich zur Baselinephase herangezogen, ansonsten andere Responsemaße einschl. Präferenzangaben bei cross-over Studien (siehe unten). Ursprünglich war vorgesehen, den Responderanteil in den einzelnen Gruppen nach dem intent-to-treat Prinzip zu berechnen (Zahl Responder/Zahl in die Gruppe randomisierter Patienten). Da die Zahlenangaben in vielen Studien sehr unklar waren, musste die Berechnung auf der Basis der jeweils in der Studie präsentierten Zahlen ausgewerteter Patienten durchgeführt werden, bei der Studienabbrucher in der Regel nicht berücksichtigt waren. Die berichteten Daten zur Häufigkeit variierten stark: Je nach Studie waren Attackenzahl, Zahl der Kopfschmerztag, der absolute oder der prozentuale Rückgang einer dieser beiden Maße gegenüber der Baselinephase angegeben. Außerdem variierten die Bezugszeiträume (meist 4 Wochen, teilweise aber auch 1, 8 oder mehr Wochen). Für die Effektmaßberechnung stellen jeweils niedrigere Werte günstigere Resultate (geringere Häufigkeit) dar. Dementsprechend wurden Rückgangswerte mit einem negativen Vorzeichen versehen. Zur Berechnung eines Effektmaßes mußten für die jeweilige Studie Mittelwerte und Standardabweichung oder Standardfehler (Rückberechnung der Standardabweichung durch Multiplikation mit \sqrt{n}) vorhanden sein. Studien, bei denen Mediane und Quartile oder ausschließlich Mittelwerte angegeben waren,

konnten nicht berücksichtigt werden. Da manche Studien nur die Gesamtfallzahl, nicht jedoch die Fallzahl in den einzelnen Gruppen angaben, wurde in diesen Fällen angenommen, dass jeweils die Hälfte der Patienten in einer Gruppe behandelt wurde.

Als Effektmaß bezüglich Therapieerfolg (Response) wurde das relative Risiko bzw. eine Responderratio $((\text{Zahl der Responder in Testgruppe} / \text{Zahl der Patienten in Testgruppe}) / (\text{Zahl der Responder in Kontrollgruppe} / \text{Zahl der Patienten in Kontrollgruppe}))$ mit dem zugehörigen 95%-Konfidenzintervall berechnet. Werte > 1 weisen auf eine Überlegenheit der Prüfgruppe, Werte < 1 auf eine Überlegenheit der Kontrollgruppe. Für die Häufigkeitsmaße wurden standardisierte Mittelwertsdifferenzen $(\text{Mittelwertsdifferenz} / \text{gepoolte Standardabweichung})$ und 95%-Konfidenzintervalle berechnet. Werte > 0 weisen auf eine Überlegenheit der Testgruppe.

Als Effektmaß bzgl. Nebenwirkungen wurde das relative Risiko $((\text{Zahl der Patienten mit Nebenwirkungen in Testgruppe} / \text{Zahl der Patienten in Testgruppe}) / (\text{Zahl der Patienten mit Nebenwirkungen in Kontrollgruppe} / \text{Zahl der Patienten in Kontrollgruppe}))$ mit dem zugehörigen 95%-Konfidenzintervall berechnet. Werte < 1 deuten hier auf ein günstigeres Ergebnis für die Testgruppe.

In der Cochrane Collaboration wird häufig empfohlen, bei cross-over Studien zur Effektmaßberechnung für Meta-Analysen nur Daten für die erste (parallele) Behandlungsphase vor dem cross-over heranzuziehen. Da mehr als 40% der Studien ein cross-over Design hatten, jedoch nur einzelne dieser Studien Daten für die erste Phase separat präsentierten, wäre eine Effektmaßberechnung dementsprechend für diese Studien nicht möglich. Da von der Berechnung gepoolter Effektgrößen abgesehen (siehe nächster Absatz) und der Informationsverlust durch den Ausschluss dieser großen Anzahl von Studien als äußerst ungünstig angesehen wurde, erfolgte auch für cross-over Studien eine Effektmaßberechnung, wobei die Ergebnisse beider Phasen herangezogen wurden.

Die Effektmaße wurden je nach Art der Kontrollgruppe (Placebo und andere Behandlung mit Untergruppen Calcium-Antagonisten, β -Blocker, andere Arzneimittel) gruppiert. Es erfolgten separate Zusammenstellungen ohne und mit cross-over Studien. Im Ergebnisteil werden in aller Regel Darstellungen mit Einschluss der cross-over Studien präsentiert, da hierfür mehr Studien Effektmaße beigetragen haben. Nur wenn der Einbezug von cross-over Studien keine zusätzliche Information erbrachte, wird die Zusammenstellung der Studien mit Paralleldesign präsentiert. Im Protokoll war festgelegt, dass ein gepooltes Effektmaß nur dann berechnet werden sollte, wenn a) zumindest für 50% der Studien, die den jeweiligen Parameter zu diesem Zeitpunkt bestimmt hatten, verwertbare Daten vorlagen und b) die Studien aus klinischer Sicht ausreichend vergleichbar waren. Angesichts der oben aufgeführten Schwierigkeiten, wurde die Berechnung gepoolter Effektmaße als obsolet angesehen.

Alle Zusammenstellungen und Berechnungen erfolgten mit Hilfe der Cochrane Collaboration Review Manager Software, Version 4.1. Aus diesem Programm stammen auch die tabellarisch-graphischen Ergebniszusammenstellungen.

Angesichts der Problematik der Effektstärkeberechnungen erfolgte entsprechend Studienprotokoll zusätzlich ein 5-stufiger vote count (+ = Test signifikant und/oder klinisch relevant besser als Kontrolle, (+) = Test im Trend (klinisch und/oder statistisch) besser als Kontrolle, 0 = kein Unterschied, (-) = Kontrolle im Trend besser als Test, - = Kontrolle signifikant und oder/klinisch relevant besser als Test (siehe auch Beschreibung in (321)). Außerdem erfolgte eine deskriptive Zusammenfassung der Ergebnisse der jeweiligen Studie.

2.2.2 Behandlung von Depressionen mit Hypericumextrakten

2.2.2.1 Ein- und Ausschlusskriterien

Folgende Bedingungen mussten für den Einschluss einer Studie erfüllt sein:

1. Patienten: Patienten mit Depressionen bzw. mit depressiven Verstimmungen
2. Intervention: In der Prüfgruppe Hypericummonoextrakte, in der Kontroll- bzw. Vergleichsgruppe Placebo oder andere Antidepressiva über mindestens 4 Wochen (im Gegensatz zu früheren Versionen der Übersichtsarbeit wurden Studien zu Kombinationspräparaten ebensowenig berücksichtigt wie Studien, in denen die Prüfpräparate weniger als 4 Wochen verabreicht worden waren)
3. Design: explizit randomisierte und quasirandomisierte (z.B. Zuteilung entsprechend Geburtsdatum) klinische Studien sowie doppelblinde Studien, in denen die aufnehmende und behandelnde Person verblindet war und daher zumindest von einer quasirandomisierten Studie auszugehen ist;
4. Zielkriterien: Messung von Zielparametern, die eine Beurteilung der Veränderung des klinischen Zustandes bzw. der depressiven Symptomatik erlauben (z.B. Depressionsskalen, Respondermaße etc.). Studien, in denen nur physiologische Parameter bestimmt wurden, fanden keine Berücksichtigung.

2.2.2.2 Literatursuche

Potentiell relevante Studien wurden in folgenden Datenbanken bzw. auf folgende Weise gesucht:

- Datenbank der Cochrane Depression, Anxiety and Neurosis Group (für diese Datenbank werden regelmäßig eine Reihe von Datenbanken wie Medline, Embase, Pschlit, Psychindex etc. durchsucht, außerdem Handsuchen in psychiatrischen Zeitschriften und Kongressbänden durchgeführt) (347);
- Zusatzsuchen in PubMed (Suchbegriff hyperic*), zuletzt im September 2001;
- Durchsicht der Literaturverzeichnisse von gefundenen Studien und Übersichtsarbeiten;
- Anfragen bei Herstellern und Experten;
- Durchsicht von Kongressbänden von Phytotherapietagen.

Eine Einschränkung der Publikationssprache gab es nicht.

2.2.2.3 Selektionsprozess

In einem ersten Screeningschritt wurden von einem Reviewer alle Referenzen aussortiert, bei denen eindeutig erkennbar war, daß sie die Einschlusskriterien nicht erfüllen (unkontrollierte Studien, Reviews, Kasuistiken, pharmakokinetische Studien). Alle potentiell relevanten Referenzen wurden im Volltext besorgt und formal von mindestens zwei Beurteilern mit einem Standardbogen auf die Erfüllung der Einschlusskriterien geprüft. Die Ausschlussgründe wurden jeweils dokumentiert.

2.2.2.4 Extraktion und Beurteilung

Die Extraktion von bibliographischen Angaben, Information zu Patienten, Methodik, Ergebnissen und Schlußfolgerungen sowie die Beurteilung der Studienqualität erfolgte mit einem vorgeprüften Formular durch mindestens zwei Reviewer. Ergebnisdaten wurden nach Möglichkeit zu folgenden Parametern extrahiert: Summenscorewerte entsprechend Hamilton Rating Scale for Depression (vor Behandlung, 1-2, 3-4, 5-6 sowie >6 Wochen nach Therapiebeginn und nach Therapieende allgemein), Anzahl Responder bei Therapieende (mit Dokumentation der Responderdefinition), Werte für andere Bewertungsskalen (vor Therapie und bei Therapieende), Anzahl der Patienten mit Nebenwirkungen, Anzahl der Studienabbrüche (insgesamt sowie wegen Nebenwirkungen, mangelnder Wirksamkeit und anderen Gründen).

Die Bewertung der *methodischen Qualität* erfolgte mit Hilfe des Jadadscores (für detaillierte Angaben siehe Abschnitt 2.2.1.4.) sowie mithilfe einer vom Autor entwickelten Skala (Internal

Validity Scale). Diese Skala war bei der ersten Bearbeitung der bis 1995 vorliegenden Studien benutzt worden, weil der Jadascore noch nicht offiziell verfügbar war. Genauere Angaben zur Internal Validity Scale finden sich in Abschnitt 3.2.2. Für die hier präsentierte Auswertung wird jedoch nur das auch in der Delphiliste (siehe 2.2.1.4.) enthaltene Item zur Verblindung der randomisierten Zuteilung (concealment) herangezogen. Als adäquate Concealmentmethoden werden dabei konsekutiv numerierte, neutral verpackte Medikationspackungen oder eine zentrale Randomisation per Telefon und äquivalente Methoden angesehen.

2.2.2.5 Zusammenfassung der Primärstudienresultate

Die Mehrzahl der eingeschlossenen Studien beurteilte den klinischen Verlauf der depressiven Symptomatik mit Hilfe der Fremdbeurteilungsskala Hamilton Rating Scale für Depression (HAMD). Weitere häufig benutzte Instrumente waren der Clinical Global Impression Index (CGI, ebenfalls Beurteilung durch den Arzt) sowie die Selbstbeurteilungsinstrumente SDS (Self Rating Depression Scale, Zung) und D-S (Depressions-Skala, von Zerssen). Fast alle Studien präsentierten ausserdem ein Respondermaß. Häufigstes Responderkriterium war ein Rückgang des Scorewertes auf der HAMD um mindestens 50% gegenüber dem Ausgangswert bzw. auf einen Scorewert von maximal 10. Mehrfach war außerdem als Responderkriterium ein mindestens „deutlich gebesserter“ Zustand entsprechend CGI angegeben. Gelegentlich gab es Globalurteile bzgl. des Therapieerfolgs von Arzt oder Patient.

Hauptzielkriterium zur Beurteilung der Wirksamkeit war das relative Risiko, hier wegen der Missverständlichkeit des Begriffs bei einem positiven Therapieergebnis als Responderratio ((Zahl der Responder unter Hypericum/Zahl der mit Hypericum behandelten Patienten)/(Zahl der Responder in der Vergleichsgruppe/Zahl der in der Vergleichsgruppe behandelten Patienten)) bezeichnet, mit dem entsprechenden 95%-Konfidenzintervall bei Therapieende. Nach Möglichkeit wurde das HAMD-Responsekriterium herangezogen, in zweiter Präferenz die obengenannte CGI-Beurteilung und als dritte Präferenz andere Globalurteile. Außerdem wurden standardisierte Mittelwertsdifferenzen für HAMD-Scorewerte nach 1 bis 2, 3 bis 4, 5 bis 6 Wochen, nach mehr als 6 Wochen sowie nach dem jeweiligen Therapieende berechnet. Auch für D-S-Scorewerte wurden standardisierte Mittelwertsdifferenzen nach Therapieende berechnet.

Hauptzielkriterium für die Beurteilung der Nebenwirkungen (nur im Vergleich zu konventionellen Antidepressiva) war die Zahl der Patienten mit Nebenwirkungen. Berechnet wurde wiederum das relative Risiko und das entsprechende 95%-Konfidenzintervall. Nebenzielvariablen waren die Gesamtzahl von Studienabbrüchen sowie die Zahl von Patienten, die wegen Nebenwirkungen die Studien abbrachen.

Für die weitere Auswertung wurden Studien zu Hypericumextrakten vs. Placebo und zu Hypericumextrakten vs. konventionelle Antidepressiva strikt getrennt. Innerhalb dieser Vergleiche

wurden die Extrakte unterschieden (die zwei am häufigsten geprüften Extrakte jeweils separat und die übrigen Extrakte gemeinsam). Die integrierenden statistischen Analysen (Meta-Analysen) erfolgten, soweit nicht anders angegeben, auf intent-to-treat Basis (d.h. als Anzahl der behandelten Patienten wurde die in die jeweilige Gruppe randomisierte Zahl von Patienten herangezogen). Zusätzliche Analysen erfolgten (a) für die Studien, die explizit nur Patienten mit majorer Depression einschlossen, (b) ausschließlich für Studien mit HAMD-Responsekriterium und (c) für den Responder-Vergleich von Hypericumextrakten und konventionellen Antidepressiva nach dem per protocol Prinzip (nur Patienten, die die Studien protokollgemäß abschlossen). Funnel plot-Graphiken wurden für die beiden Hauptvergleiche zur Wirksamkeit erstellt. Sensitivitätsanalysen wurden bzgl. des Einflusses des Publikationsjahres und dem Schweregrad der Depressionen (mittlerer HAMD-Score vor Behandlungsbeginn) durchgeführt. Alle Berechnungen erfolgten mit der Cochrane Collaboration Review Manager Software 4.1.

2.3 Ergebnisse

2.3.1 Migräneprophylaxe mit Propranolol, Metoprolol und Flunarizin

2.3.1.1 Literatursuche und Selektion

Zu Propranolol wurden durch den Literatursuche- und Screeningprozess insgesamt 93 potentiell relevante Publikationen identifiziert. 57 Studien, publiziert in 55 Hauptveröffentlichungen (6;9;10;18;26;35;44;45;47;78;82;83;86;88;134;135;144;146;147;153;178;187;218;227-230;237;238;257;297;298;304;306;310;325;339;343;350;353;357;370;379;392;393;402;409;418;436;438;445;459;492;506;530) und 14 Zweitveröffentlichungen (46;48;90;143;346;378;380;395;424-426;431;437;531) entsprachen den Einschlusskriterien. 25 Publikationen entsprachen nicht den Einschlusskriterien: 6 waren Übersichtsarbeiten (1;14;334;384;458;464), 10 berichteten nichtrandomisierte klinische Studien (75;79;81;222;388;397;432;471;472;519), eine Studie war nicht verfügbar (41), 2 waren Studien zur Behandlung akuter Attacken (27;140), eine hatte eine Beobachtungsdauer von weniger als 4 Wochen (517) und in 5 randomisierten Studien war die Kontrollintervention weder Placebo noch eine andere medikamentöse Behandlung (58;172;188;363;423).

28 Publikationen wurden zu Metoprolol identifiziert. Insgesamt 16 Studien erfüllten die Einschlusskriterien (17;89;146;154;155;177;226;227;262;295;353;402;422;435;477;523); zu 4 Studien lag eine Zweitveröffentlichung vor (90;157;178;524). 8 Publikationen wurden ausgeschlossen: 2 waren Übersichtsarbeiten (384;458), 3 berichteten nichtrandomisierte klinische

Studien (397;432;519), jeweils eine Studie hatte eine nicht den Einschlusskriterien entsprechende Kontrollintervention (181), Dauer (517) oder Patientenstichprobe (394).

Zu Flunarizin wurden insgesamt 73 potentiell relevante Publikationen gefunden. 40 Studien, publiziert in 39 Hauptveröffentlichungen erfüllten die Einschlusskriterien (3;8;13;25;44;45;54;63-65;80;86;97;138;144;145;155;156;261;263;264;294;296-298;323;326;341;345;362;366;382;402;409;421;422;425;433;460;522). In einem Fall wurde eine Übersichtsarbeit eingeschlossen, die Daten zu einer ansonsten unveröffentlichten Studie berichtete (13). Zweitpublikationen lagen zu 10 Studien vor (6;136;137;143;157;163;367;424-426). 24 Publikationen wurden ausgeschlossen: 9 waren Übersichtsarbeiten (14;15;334;351;384;397;417;428;446), 2 nicht verfügbar (16;344), 9 waren nichtrandomisierte Studien (43;77;79;307;309;324;471;472;519), in einer Studie waren Patienten mit verschiedenen Kopfschmerzdiagnosen gemeinsam ausgewertet worden (55), in einer weiteren wurde Flunarizin nicht selbst geprüft (134) und in 2 Studien entsprach die Art der Kontrollintervention nicht den Einschlusskriterien (12;30). Unter Berücksichtigung der Überschneidungen wurden zu Propranolol, Metoprolol und Flunarizin insgesamt 97 Studien in die Bewertung eingeschlossen.

2.3.1.2 Allgemeine Studienmerkmale

Wichtige Informationen zu den allgemeinen Studienmerkmalen aller Studien zur Migräneprophylaxe mit Propranolol, Metoprolol und Flunarizin sind in Tabelle 1 zusammengestellt. Zusammenfassende Informationen zu Patienten, Methoden und Ergebnissen der einzelnen Studien finden sich in den Tabellen A1 – A10. Aufgrund ihres großen Umfangs werden diese Tabellen als Anhang am Ende dieser Monographie präsentiert.

25 Propranololstudien waren placebokontrolliert, 41 verglichen Propranolol mit anderen Medikamenten (in 9 Studien wurde sowohl mit Placebo wie auch mit anderen Medikamenten verglichen). Für Metoprolol lauten die entsprechenden Zahlen 4 und 13, für Flunarizin 10 und 31 Studien. Die Studien wurden zwischen 1972 und 2001 meist in englischer Sprache publiziert. Nur bei 14 Studien war angegeben, wie die Studie finanziert wurde. Die mediane Zahl randomisierter Patienten über alle 97 Studien betrug 45 (Bereich 9-783); in Studien, in denen zwei Medikamente verglichen wurden, war die mediane Patientenzahl 44, in Studien die sowohl eine Placebo- wie auch eine aktive Kontrolle hatten, 59 Patienten. Soweit dies nachvollziehbar war, wurden im Schnitt 17% der randomisierten Patienten nicht in die Auswertung einbezogen; die Zahl der tatsächlichen Studienabbrüche liegt höher, da in einigen Studien zumindest ein Teil der Patienten, die die Studie abbrachen, in die Auswertung einbezogen wurden. Die Dosierung lag für Propranolol zwischen 80 und 320 mg, für Metoprolol zwischen 50 und 200 mg und für Flunarizin zwischen 5 und 20 mg, meist jedoch bei 10 mg.

In 82% der Studien wurden Patienten mit Migräne mit oder ohne Aura eingeschlossen, nur in wenigen Studien erfolgte eine Beschränkung auf einen Migränetyp. 44 Studien (45%) hatten ein cross-over Design; ältere Studien nutzten dieses Design häufiger als neuere. Die Dauer der Behandlungsphase lag in den meisten Studien zwischen 8 und 16 Wochen. Nur wenige Studien beobachteten die Patienten auch nach Abschluss der Behandlung weiter. Die Mehrzahl dieser Studien hatte wiederum hohe Drop-out Raten.

Tabelle 1:
Randomisierte Studien zur Migräneprophylaxe mit Propranolol, Metoprolol und Flunarizin: Allgemeine Studienmerkmale

	Propranolol	Metoprolol	Flunarizin	Alle
Anzahl Studien				
- Insgesamt	57*	16*	40*	97*
- vs. Placebo	25 (44%)	4 (19%)	10 (25%)	39 (40%)
- vs. andere Medikamente	41 (72%)	13 (81%)	31 (78%)	68 (70%)
Publikationsjahr (Median, Bereich)	1986 (1972-1999)	1987 (1983-2001)	1988 (1980-1999)	1986 (1972-2001)
Nur als Abstract publiziert/unpubliziert	5 (9%)	-	6 (15%)	10 (10%)
In englischer Sprache verfügbar	55 (97%)	16 (100%)	32 (80%)	87 (90%)
Sponsor angegeben	9 (16%)	3 (19%)	4 (10%)	14 (14%)
Patientenzahl (Median, Bereich)	49 (9-783)	59 (28-270)	50 (27-783)	45** (9-783)
% Abbrüche/Ausschlüsse (Median, Bereich)	20% (0-50%)	15% (4-43%)	16% (0-48%)	17% (0-50%)
Studien mit Diagnose				
- Migräne (mit oder ohne Aura bzw. unklar)	49 (86%)	13 (81%)	30 (75%)	79 (82%)
- Ausschl. Migräne ohne Aura	7 (12%)	1 (6%)	8 (20%)	14 (14%)
- Ausschl. Migräne mit Aura	1 (2%)	2 (13%)	2 (5%)	4 (4%)
Kopfschmerzklassifikation				
- International Headache Society	11 (19%)	5 (31%)	7 (18%)	18 (19%)
- Ad-hoc Committee	20 (35%)	3 (19%)	17 (42%)	37 (38%)
- Andere	9 (16%)	7 (44%)	4 (10%)	16 (16%)
- Unklar/Keine	17 (30%)	1 (6%)	12 (30%)	26 (27%)
Design				
- Parallel	24 (42%)	6 (38%)	35 (88%)	53 (55%)
- Cross-over	33 (58%)	10 (62%)	5 (12%)	44 (45%)
Dauer in Wochen (Median, Bereiche)	4 (0-10)	4 (3-8)	4 (0-24)	4 (0-24)
Behandlungsphase***	12 (4-30)	8 (4-30)	14 (8-24)	12 (4-30)
Follow-up-Phase	0 (0-20)	0 (0-12)	0 (0-24)	0 (0-24)

*Mehrere Studien zu mehr als einem der drei Medikamente und zum Teil mit mehr als 2 Vergleichsgruppen

**Der Medianwert liegt in der Gesamtgruppe niedriger als in den einzelnen Subgruppen (Simpson'sches Paradox)

***Bei cross-over-Studien wurde nur die Behandlungsphase pro Medikament gezählt

2.3.1.3 Studienqualität

Sowohl die methodische Qualität wie auch die Angemessenheit der Beobachtung und Berichterstattung waren bei der Mehrzahl der Studien wenig zufriedenstellend (siehe Tabelle 2). Eindeutiges methodisches Hauptproblem der Studien war die Beschreibung von Studienabbrüchen und –ausschlüssen sowie deren Berücksichtigung in der statistischen Auswertung. Immerhin noch in 41% der Studien wurden die Anzahl und die Gründe für Studienabbrüche und –ausschlüsse

genannt, aber nur in 5 Studien wurde eine angemessene intent-to-treat Analyse durchgeführt. Obwohl eine solche Analyse bei Äquivalenzstudien nicht als Hauptanalyse empfohlen wird, ist sie doch zumindest als zusätzliche Analyse angesichts der meist hohen Drop-out Raten wünschenswert.

Tabelle 2:
Randomisierte Studien zur Migräneprophylaxe mit Propranolol, Metoprolol und Flunarizin: Qualitätsaspekte

	Propranolol	Metoprolol	Flunarizin	Alle
Jadadscore				
Randomisation				
Genannt	53 (93%)	15 (94%)	32 (80%)	85 (88%)
Randomisationsgenerierung beschrieben	1 (2%)	-	2 (5%)	3 (3%)
Doppelblindbedingungen				
Genannt	49 (86%)	16 (100%)	34 (85%)	84 (87%)
Doppelblindbed. gut/glaubhaft beschrieben	13 (23%)	8 (50%)	10 (25%)	26 (27%)
Drop-outs/Abbrüche beschrieben	24 (42%)	9 (56%)	12 (30%)	40 (41%)
Summenscore (Median, Bereich)	2 (1-4)	3 (2-4)	2 (1-5)	2 (1-5)
Delphiliste				
Explizit randomisiert	51 (90%)	15 (94%)	32 (80%)	83 (86%)
Concealment beschrieben	1 (2%)	-	1 (3%)	2 (2%)
Vergleichbarkeit vor Ther. nachvollziehbar	9 (16%)	6 (38%)	10 (25%)	21 (22%)
Einschlusskriterien spezifiziert	32 (56%)	14 (88%)	25 (63%)	60 (62%)
Beurteiler blind	48 (84%)	15 (94%)	32 (80%)	82 (85%)
Therapeut blind	49 (86%)	15 (94%)	32 (80%)	82 (85%)
Patient blind	52 (91%)	16 (100%)	34 (85%)	84 (87%)
Detaillierte Ergebnisse zu Hauptzielkriterien	37 (65%)	15 (94%)	28 (70%)	69 (71%)
Intent-to-treat Analyse	2 (4%)	2 (13%)	1 (3%)	5 (5%)
Kriterien erfüllt (Median, Bereich)	5 (1-8)	6 (0-8)	5 (1-7)	5 (0-8)
Angemessenheit Beobachtung und Bericht				
Rekrutierung beschrieben	8 (14%)	5 (31%)	8 (20%)	18 (19%)
Klare Kopfschmerzdiagnose	36 (63%)	15 (94%)	26 (65%)	66 (68%)
Patienten ausreichend charakterisiert	17 (30%)	11 (69%)	17 (43%)	38 (39%)
Mindestens 4 Wochen Baseline	33 (58%)	15 (94%)	28 (70%)	66 (68%)
Kointerventionen beschrieben	7 (12%)	6 (38%)	3 (8%)	13 (13%)
Verwendung von Kopfschmerztagebüchern	45 (79%)	15 (94%)	29 (73%)	77 (79%)
Daten zu Häufigkeit detailliert präsentiert	28 (49%)	13 (81%)	26 (65%)	59 (61%)
Daten zur Schmerzintensität detailliert präs.	14 (25%)	5 (31%)	14 (35%)	28 (29%)
≥ 90% der Pat. mit Daten über 2 Monate	10 (18%)	4 (25%)	12 (30%)	22 (23%)
≥ 80% der Pat. mit Daten über 6 Monate	3 (5%)	-	4 (10%)	5 (5%)
Kriterien erfüllt (Median, Bereich)	3 (0-8)	6 (2-8)	3 (0-8)	4 (0-8)

Nur in 22 Studien (23%) konnten nachvollziehbar Daten zu mehr als 90% der randomisierten Patienten über 2 Monate nach Therapiebeginn präsentiert werden. Daten zu mehr als 80% der randomisierten Patienten nach 6 Monaten berichteten nur 5 Studien (5%).

Die Fallzahl war insbesondere bei den Äquivalenzstudien (Median 44 Patienten) meist so gering, dass ein möglicher Unterschied zwischen den verglichenen Medikamenten nicht mit ausreichender

Sicherheit erfaßt werden konnte. Die Methoden zur Generierung der Randomisationssequenz und zur Verblindung der Zuteilung (concealment) wurden nur in Ausnahmefällen beschrieben. Auch wenn bei Medikamentenstudien in der Regel davon auszugehen ist, dass das Concealment durch fortlaufend nummerierte, neutrale Medikamentenverpackungen erfolgt, ist die Nichtbeschreibung dieses in der methodischen Literatur als Kernkriterium angesehenen Details auch in neuen Studien bedenklich.

Positiv ist anzumerken, dass fast alle Studien doppelblind waren (eine Überprüfung des Verblindungserfolges wurde jedoch in keiner Studie beschrieben). Eine klare Kopfschmerzdiagnose nach definierten Kriterien, eine mindestens vierwöchige Baselinephase, die Verwendung eines Kopfschmerztagebuches und eine zwar heterogene, aber akzeptable Präsentation der Daten zur Kopfschmerz- oder Attackenhäufigkeit war ebenfalls in der Mehrheit der Studien gegeben.

2.3.1.4 Studienergebnisse

Aufgrund der häufig unzureichenden Darstellung in den Studien muß die Zusammenfassung der Studien unter Einbezug sowohl der Effektmaße als auch des vote counts (Kategorisierung der Ergebnisse) und der Beschreibung in den Tabellen A1 bis A10 im Anhang erfolgen. Tabelle 3 gibt einen Überblick für wieviele Studien bzw. Vergleiche Effektmaßberechnungen möglich waren. In Tabelle 4 sind die Kategorisierungen der Ergebnisse zusammengestellt.

Tabelle 3:
Anzahl der Vergleiche mit verwertbaren Daten für Effektstärkenberechnungen ohne/mit Berücksichtigung von cross-over-Studien

Substanz	Propranolol	Metoprolol	Flunarizin
(Gesamtzahl Vergleiche)	(71)	(21)	(46)
Responder bzw. Therapieerfolg	24 / 37	5 / 15	22 / 29
Attacken- oder Kopfschmerzhäufigkeit	15 / 30	5 / 9	15 / 16
Anzahl Patienten mit Nebenwirkungen	14 / 24	3 / 5	13 / 13

Tabelle 4:
Kategorisierung der Studienergebnisse (vote count)

	+	(+)	0	(-)	-	unint.
Propranolol						
• vs. Placebo (25 Vergleiche)	17	6	2			
• vs. Calciumantagonisten (15 Vergleiche)		2	12	1		
• vs. andere β -Blocker (12 Vergleiche)			7	3	1	1
• vs. sonstige (19 Vergleiche)	1	5	10	2		1
Metoprolol						
• vs. Placebo (4 Vergleiche)	3	1				
• vs. Calciumantagonisten (5 Vergleiche)	1	2	2			
• vs. andere β -Blocker (6 Vergleiche)		2	3			1
• vs. sonstige (6 Vergleiche)	5			1		
Flunarizin						
• vs. Placebo (10 Vergleiche)	6	2	1			1
• vs. andere Calciumantagonisten (9 Vergleiche)	1	2	5	1		
• vs. β -Blocker (12 Vergleiche)		1	10	1		
• vs. sonstige (15 Vergleiche)	1	5	7	1		1

+ = Test signifikant und/oder klinisch relevant besser als Kontrolle, (+) = Test im Trend (klinisch und/oder statistisch) besser als Kontrolle, 0 = kein Unterschied, (-) = Kontrolle im Trend besser als Test, - = Kontrolle signifikant und oder/klinisch relevant besser als Test, unint. = uninterpretierbar

Propranolol vs. Placebo:

Bei der Kategorisierung der Ergebnisse wurde Propranolol in insgesamt 25 Vergleichen mit Placebo in 17 Fällen als signifikant bzw. deutlich überlegen eingestuft, in 6 Fällen war ein Trend zugunsten von Propranolol gegeben und in 2 Fällen kein Unterschied zu Placebo (siehe Tabellen 3 und A1 im Anhang).

Comparison: 01 Propranolol vs. placebo

Outcome: 04 responder including cross-over trials

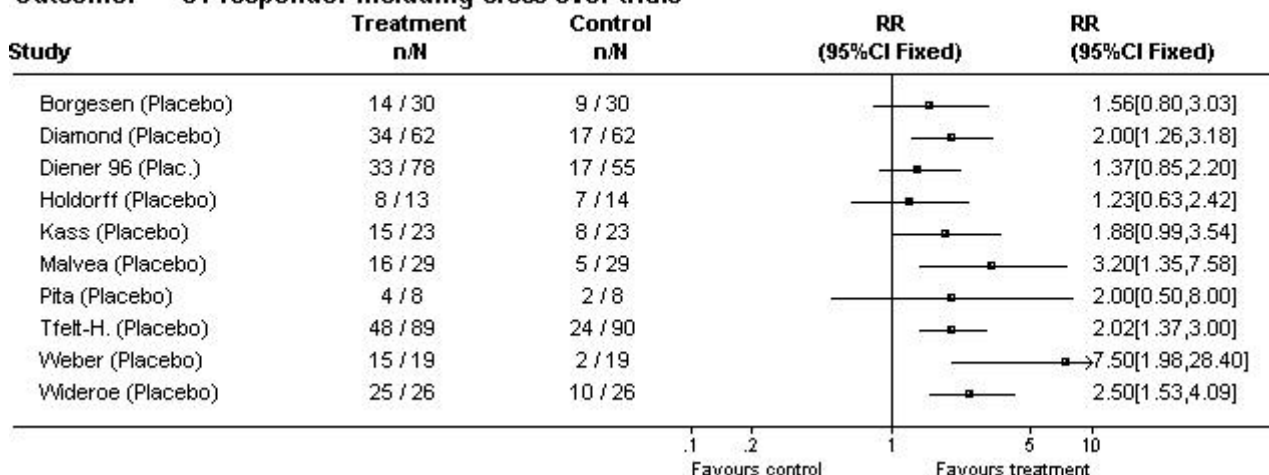


Abbildung 1:

Responderratios in Studien Propranolol vs. Placebo

(Abbildung aus Cochrane Collaboration Review Manager Software, Version 4.1; siehe Anhang zur Erläuterung)

n = Anzahl Responder, N = Anzahl Patienten in der jeweiligen Gruppe, RR = Responderratio, 95%CI fixed = 95% Konfidenzintervall

Für 10 Studien war die Berechnung von Responderratios möglich. Diese bestätigen grundsätzlich die Überlegenheit gegenüber Placebo, allerdings sind die Unterschiede in 5 Fällen nicht signifikant (siehe Abbildung 1). Bezüglich Kopfschmerzhäufigkeit präsentierten nur 2 Studien ausreichend detaillierte Ergebnisse für eine Effektmaßberechnung. In den 6 Studien, die Angaben zur Anzahl der Patienten mit Nebenwirkungen machten, war diese unter Propranolol im Schnitt etwa um ein Drittel höher als unter Placebo.

Propranolol vs. andere Substanzen:

Bei den 15 Vergleichen mit Flunarizin und anderen Calciumantagonisten ergaben sich in 12 Fällen keine eindeutigen Unterschiede, in zwei Fällen war ein Trend zugunsten von Propranolol, in einem Fall zugunsten eines Calciumantagonisten zu beobachten (siehe Tabellen 4 und A2 bis A4 im Anhang). Bei 11 interpretierbaren Vergleichen mit Metoprolol und anderen β -Blockern ergab sich in 7 Fällen kein Unterschied; zwei Studien fanden unter Metoprolol und eine unter Nadolol etwas bessere Ergebnisse, in einer Studie ergaben sich unter 160 mg Nadolol deutlich bessere Resultate als unter 160 mg Propranolol. Insgesamt 19 Vergleiche erfolgten mit einer Reihe weiterer Substanzen (Femoxetin, Acetylsalicylsäure, Methysergid, Dihydroergotamin, Cyclandelat, Mefenaminsäure, Tolfenaminsäure, Valproinsäure, Hydroxytryptophan, Amitriptylin, Clonidin und Naproxen). Eine eindeutige Überlegenheit ergab sich nur gegenüber Amitriptylin bei Patienten, die explizit keine zusätzlichen Spannungskopfschmerzen hatten. Bei beiden Vergleichen mit Ergotaminpräparaten war Propranolol zumindest tendenziell überlegen. Bei den übrigen Vergleichen war die Wirksamkeit meist ähnlich. Angesichts der häufig sehr kleinen Fallzahlen sind alle Vergleiche mit anderen Medikamenten ohne einen signifikanten Unterschied jedoch mit großer Zurückhaltung zu interpretieren.

In Abbildung 2 sind die 27 Vergleiche von Propranolol und anderen Substanzen bzgl. Responderratios zusammengefasst. Die Ergebnisse bestätigen diejenigen der Kategorisierung. Signifikante Unterschiede sind nur bei einer einzigen Studie zu erkennen.

Abbildung 3 zeigt die Ergebnisse von 18 Studien mit verwertbaren Angaben zur Häufigkeit. Hier zeigt sich allerdings nur in drei Studien kein Unterschied zu anderen β -Blockern. Demgegenüber ergibt sich zum Teil eine Unterlegenheit beim Vergleich mit anderen Substanzen.

Comparison: 02 Propranolol vs. other drugs

Outcome: 04 responder including cross-over trials

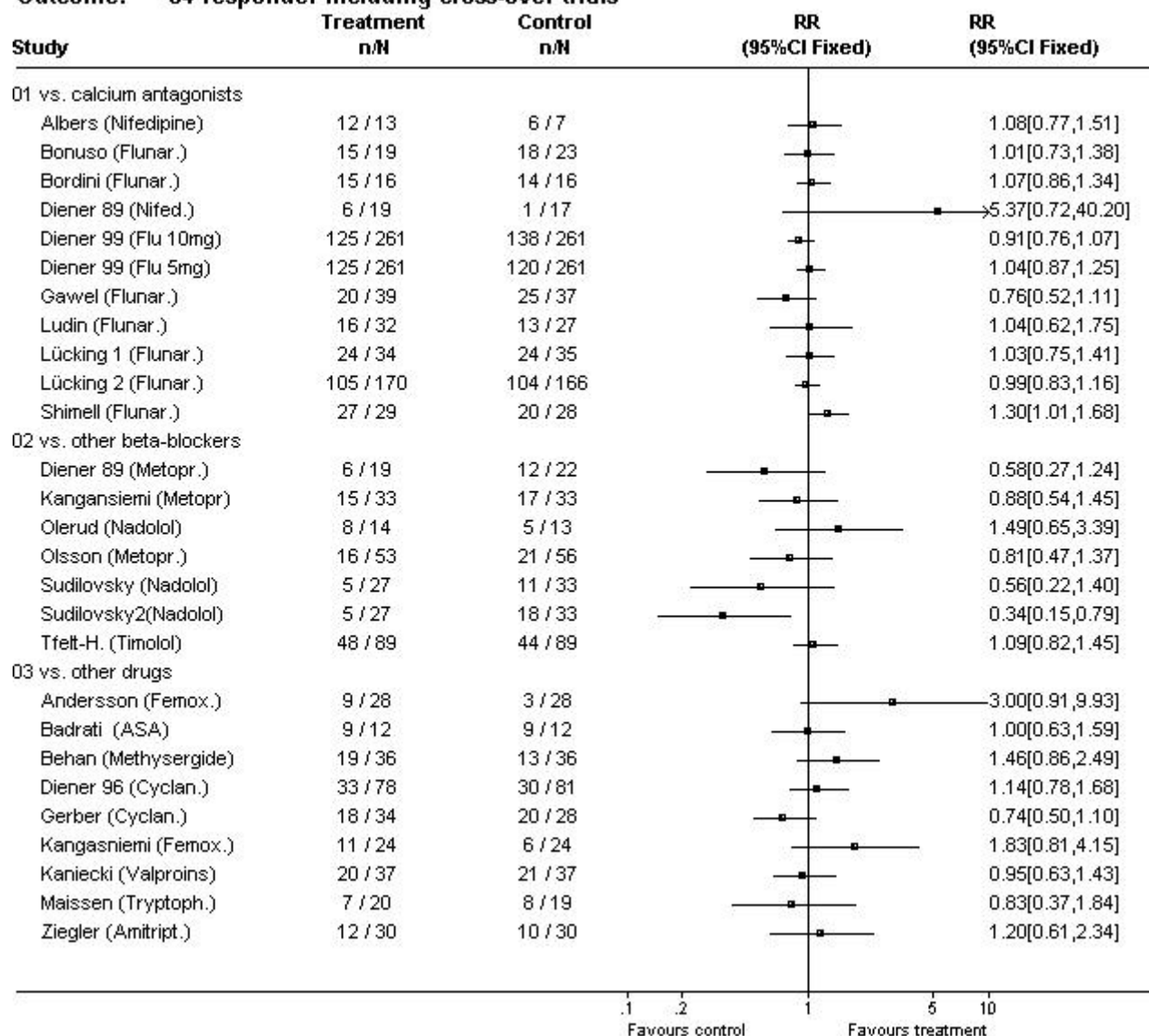


Abbildung 2:

Responderratios in Studien Propranolol vs. andere Substanzen

(Abbildung aus Cochrane Collaboration Review Manager Software, Version 4.1)

n = Anzahl Responder, N = Anzahl Patienten in der jeweiligen Gruppe, RR = Responderratio, 95%CI fixed = 95% Konfidenzintervall

Comparison: 02 Propranolol vs. other drugs
Outcome: 05 frequency measures including cross-over trials

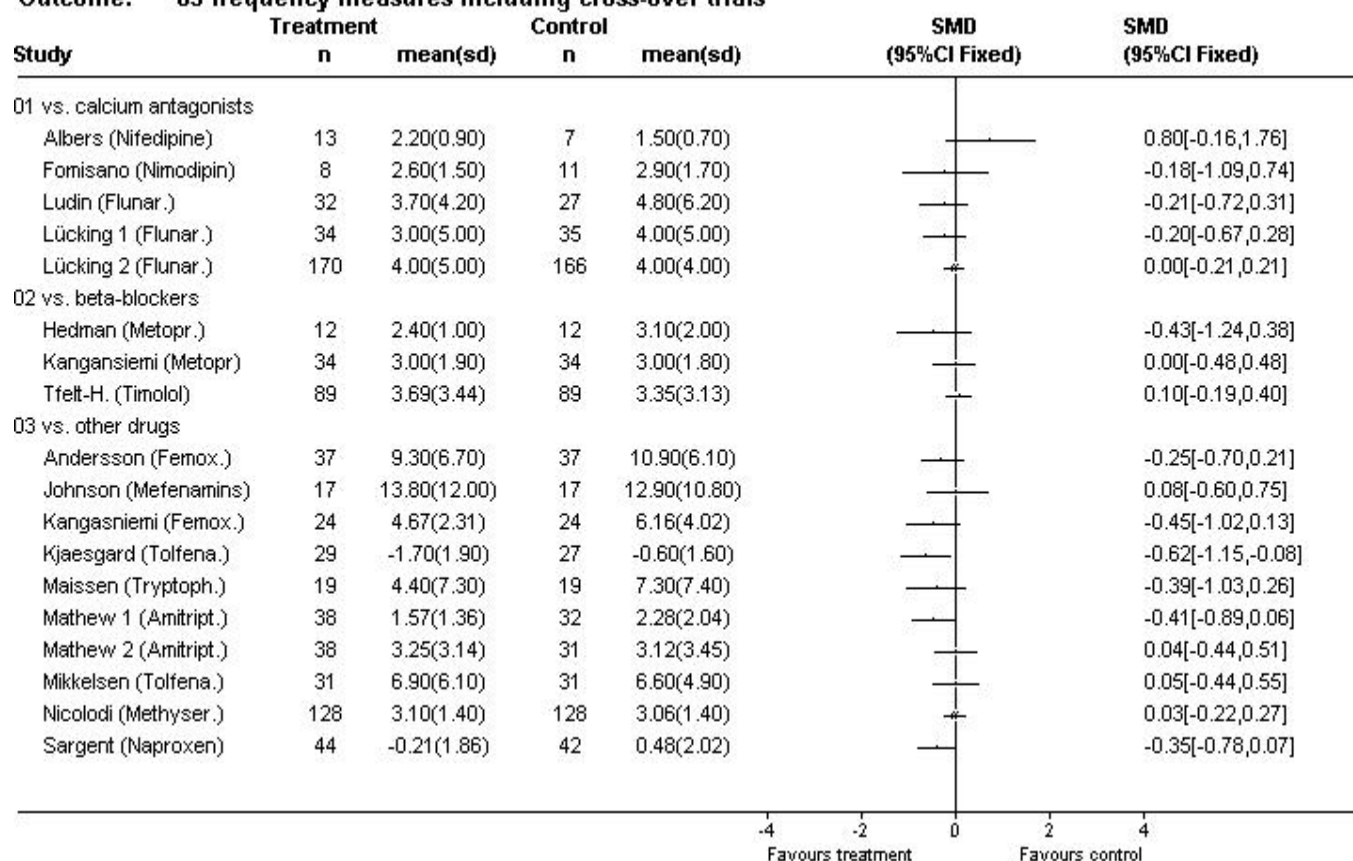


Abbildung 3:
Standardisierte Effektgrößen zu Häufigkeitsmaßen in Studien Propranolol vs. andere Substanzen
(Abbildung aus Cochrane Collaboration Review Manager Software, Version 4.1)
n = Anzahl Patienten in der jeweiligen Gruppe, mean (sd) = Mittelwert (Standardabweichung), SMD = standardisierte Mittelwertsdifferenz, 95%CI fixed = 95%-Konfidenzintervall

Bezüglich der Anzahl der Patienten mit Nebenwirkungen (siehe Abbildung 4) gibt es in den 18 Studien mit entsprechenden Daten einen leichten Trend zugunsten einer besseren Verträglichkeit von Propranolol.

Comparison: 02 Propranolol vs. other drugs

Outcome: 09 number of patients with side effects including cross-over trials

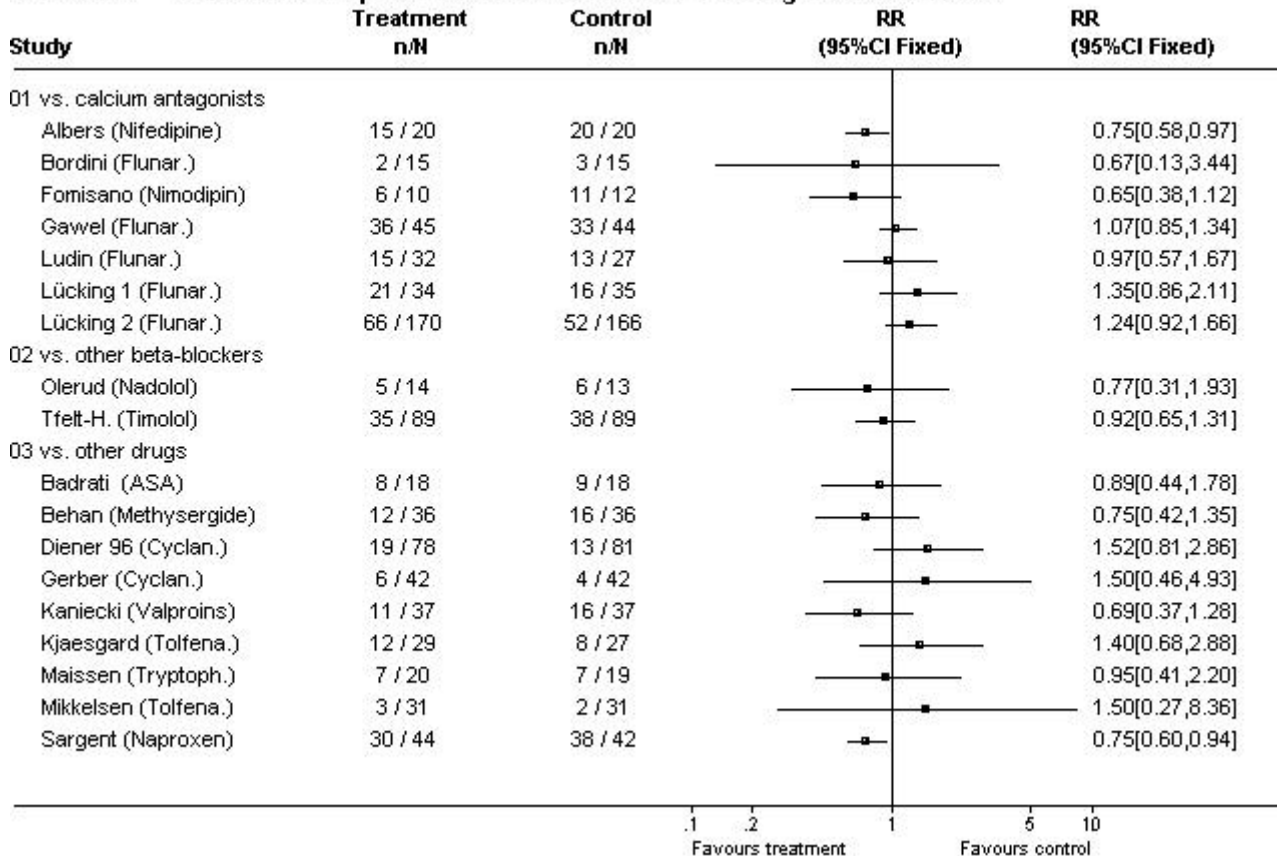


Abbildung 4:
Relative Risiken bzgl. des Auftretens von Nebenwirkungen in Studien Propranolol vs. andere Substanzen
(Abbildung aus Cochrane Collaboration Review Manager Software, Version 4.1)
n = Anzahl Responder, N = Anzahl Patienten in der jeweiligen Gruppe, RR = relatives Risiko, 95%CI fixed = 95% Konfidenzintervall

Metoprolol vs. Placebo:

Bei 4 Vergleichen mit Placebo erwies sich Metoprolol dreimal als signifikant überlegen und einmal zumindest als tendenziell überlegen (Tabelle 3 und Tabelle A5 im Anhang). Die Ergebnisse der 3 Studien mit Responderangaben und der 2 Studien mit verwertbaren Daten zur Häufigkeit bestätigen die Ergebnisse der Kategorisierung weitgehend (ohne Abbildung).

Metoprolol vs. andere Substanzen:

Auch bei Vergleichen mit anderen Medikamenten schneidet Metoprolol nur in einer Studie (im Vergleich zu Pizotifen) tendenziell schlechter ab (siehe Abbildung 5, Tabelle 3 und Tabelle A6 im Anhang). Wiederum ist bei den Vergleichen mit anderen Medikamenten die geringe Fallzahl zu beachten. Aufgrund der geringen Zahl von Studien mit Daten wurde auf eine Abbildung zu Kopfschmerzhäufigkeit und Anzahl der Patienten mit Nebenwirkungen verzichtet.

Comparison: 02 metoprolol vs. other drugs

Outcome: 04 responder including cross-over trials

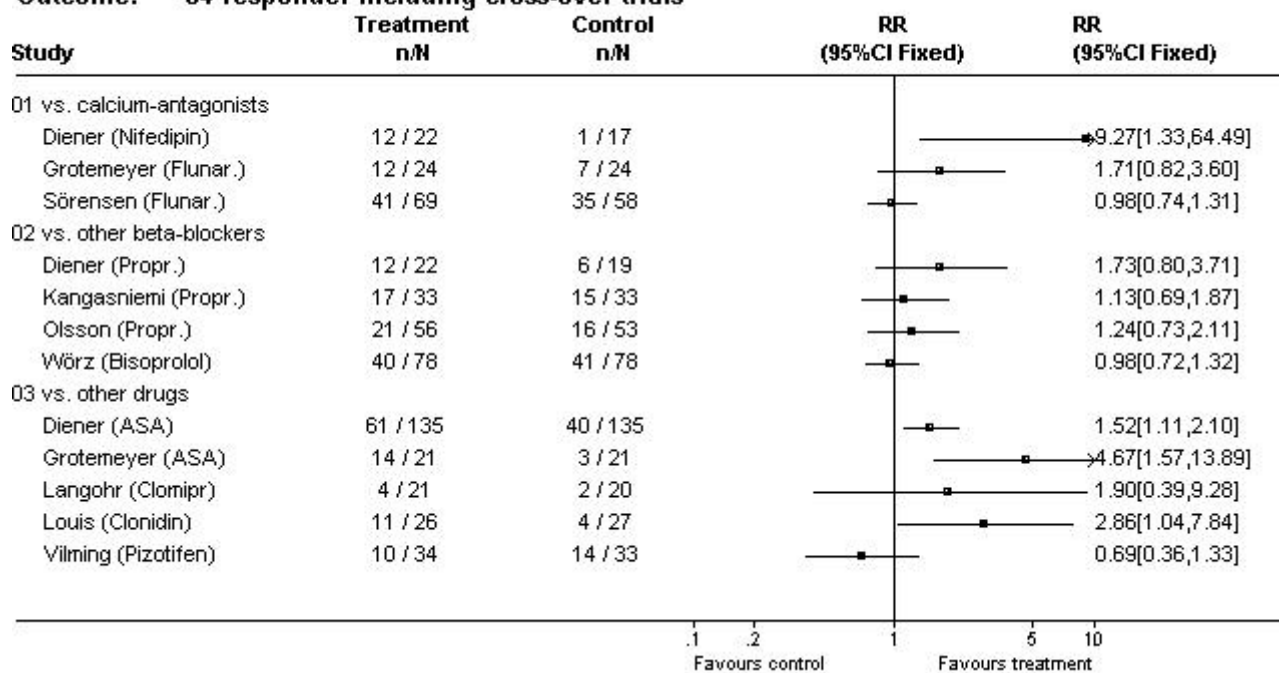


Abbildung 5:

Responderratios in Studien Metoprolol vs. andere Substanzen

(Abbildung aus Cochrane Collaboration Review Manager Software, Version 4.1)

n = Anzahl Responder, N = Anzahl Patienten in der jeweiligen Gruppe, RR = Responderratio, 95%CI fixed = 95% Konfidenzintervall

Flunarizin vs. Placebo

Flunarizin erwies sich in 9 interpretierbaren Vergleichen Placebo achtmal überlegen (6 mal signifikant, zweimal tendenziell); nur in einer Studie war kein Unterschied zu beobachten (siehe Tabelle 4 und Tabelle A7 im Anhang). Die Ergebnisse der 6 Studien, die Daten zu Responderratios präsentieren, (siehe Abbildung 6) und der 5 Studien mit Daten zur Kopfschmerzhäufigkeit (ohne Abbildung) bestätigen die Ergebnisse der Kategorisierung weitgehend, allerdings sind bei der Berechnung der Responderratio nur 2 Unterschiede signifikant. In den 4 Studien mit Daten zur Anzahl der Patienten mit Nebenwirkungen wurden keine Unterschiede im Vergleich zu Placebo berichtet.

Comparison: 01 flunarizine vs. placebo

Outcome: 04 responder including cross-over trials

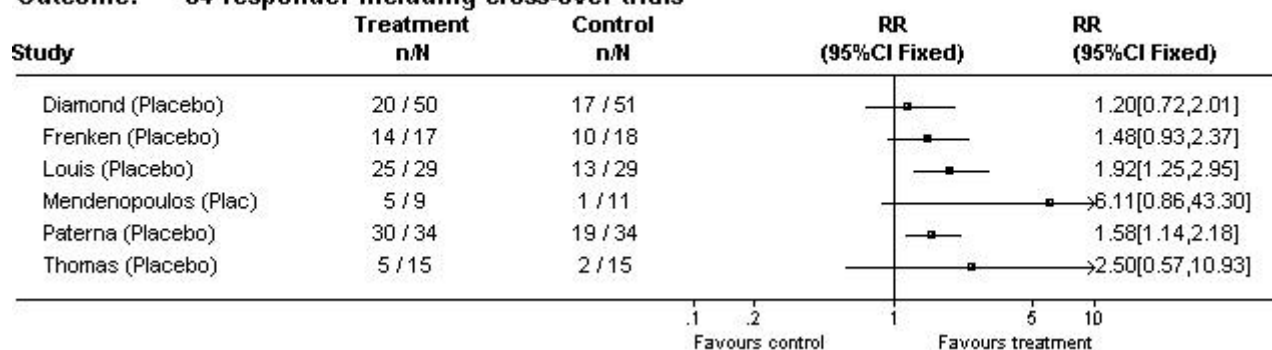


Abbildung 6: Responderratios in Studien Flunarizin vs. Placebo

(Abbildung aus Cochrane Collaboration Review Manager Software, Version 4.1)

n = Anzahl Responder, N = Anzahl Patienten in der jeweiligen Gruppe, RR = Responderratio, 95%CI fixed = 95% Konfidenzintervall

Flunarizin vs. andere Substanzen:

In den 9 Vergleichen mit anderen Calciumantagonisten erwies sich Flunarizin einmal gegenüber Nimodipin als eindeutig überlegen (beim zweiten Vergleich mit dieser Substanz war kein eindeutiger Unterschied zu erkennen). Ansonsten wurden nur geringe Unterschiede beobachtet. Dies gilt auch für die Vergleiche mit β -Blockern. In 6 Vergleichen mit Pizotifen war in drei Fällen ein tendenziell besseres Ergebnis unter Flunarizin zu beobachten, in drei Studien kein Unterschied. Beim Vergleich zu Cyclandelat ergab sich in einer Studie ein signifikant besseres Ergebnis mit Flunarizin. Ansonsten ergeben sich zu weiteren Substanzen keine überzeugenden Unterschiede. Auch hier kann aufgrund der geringen Fallzahl bei fehlenden Unterschieden aber nicht von einer Äquivalenz ausgegangen werden (siehe Tabelle 4 sowie Tabellen A8 bis A10 im Anhang).

Die Kategorisierung der Ergebnisse wird durch die Effektmaßberechnungen für die Responderangaben (siehe Abbildung 7) und für die Kopfschmerzhäufigkeitsmaße (ohne Abbildung) wiederum weitgehend bestätigt.

Im Vergleich zu Propranolol scheinen unter Flunarizin tendenziell mehr Patienten Nebenwirkungen anzugeben. Da jedoch nur sehr wenige Studien verwertbare Angaben hierzu machen, ist dieses Ergebnis mit Vorbehalt zu bewerten.

Comparison: 02 flunarizine vs. other drugs

Outcome: 04 responder including cross-over trials

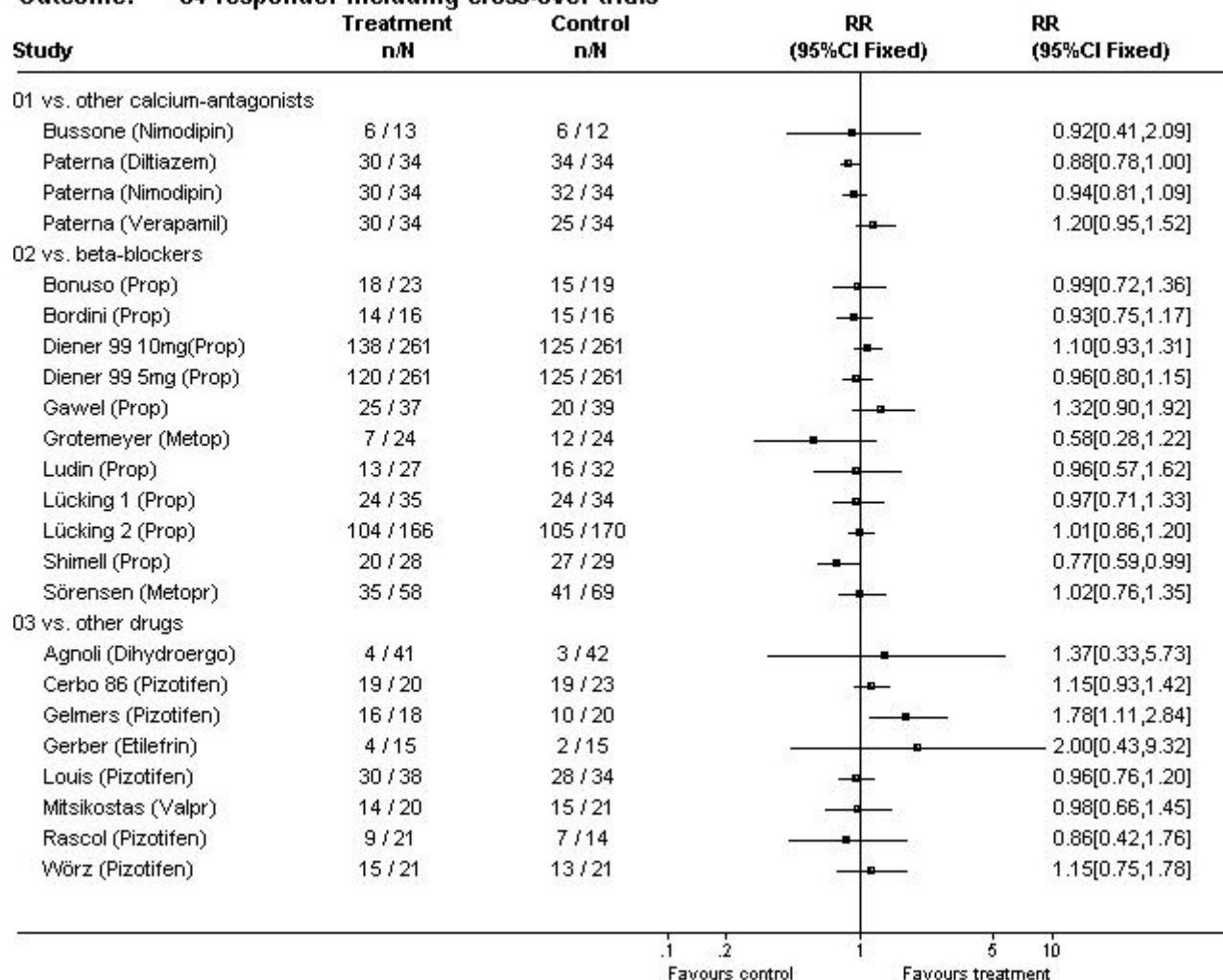


Abbildung 7:

Responderratios in Studien Flunarizin vs. andere Substanzen

(Abbildung aus Cochrane Collaboration Review Manager Software, Version 4.1)

n = Anzahl Responder, N = Anzahl Patienten in der jeweiligen Gruppe, RR = Responderratio, 95%CI fixed = 95% Konfidenzintervall

2.3.2 Behandlung von Depressionen mit Hypericumextrakten

2.3.2.1 Literatursuche und Selektion

Durch die Literatursuche wurden insgesamt 61 potenziell relevante kontrollierte klinische Studien zu Hypericumpräparaten identifiziert, die formal auf Erfüllung der Einschlusskriterien geprüft wurden. 33 Studien in 43 Publikationen entsprachen den Einschlusskriterien (38;50;159;160;164-169;173;174;186;194;196;250;254;258;259;273;335;338;354;365;381;385;396;399;400;403;404;407;419;420;484;486-488;499;515;516;518;520). 28 Studien wurden aus folgenden Gründen ausgeschlossen: in 8 Studien ging es nicht primär um die Behandlung depressiver Verstimmungen bzw. war die Einschlussdiagnose nicht Depression (11;36;91;175;193;253;303;413), 7 Studien wurden an gesunden Freiwilligen durchgeführt (52;180;216;217;398;405;430), in 5 Studien erfolgte kein Vergleich mit Placebo oder einem anderen Antidepressivum (40;276;308;427;528), 5 Studien waren zu Kombinationspräparaten mit Hypericum (92;248;358;434;507), eine Studie war nicht verfügbar (4), eine weder explizit randomisiert noch doppelblind (490) und in einer war die Beobachtungszeit unter 4 Wochen (497).

2.3.2.2 Studienmerkmale

Eine Zusammenstellung der Patientenmerkmale, Methoden, Interventionen und Ergebnisse aller eingeschlossenen Studien findet sich in den Tabellen A11 und A12 im Anhang. 23 Studien verglichen Hypericumextrakte mit Placebo, 12 mit anderen Medikamenten (2 Studien mit Placebo und aktiver Kontrolle). In 4 Studien war die Vergleichsmedikation ein selektiver Serotonin-Wiederaufnahmehemmer (dreimal Fluoxetin 20 mg täglich, einmal Sertralin 75 mg). In 7 Studien dienten zum Teil sehr niedrig dosierte, ältere Antidepressiva (viermal Imipramin in Dosierungen zwischen 75 und 150 mg, zweimal Amitriptylin 30 bzw. 75 mg und einmal Maprotilin 75 mg) und in einer älteren Studie Bromazepam als Vergleichssubstanz. Aufgrund der unterschiedlichen Klassifikationssysteme (ICD 9, ICD 10 und DSM) ist es schwer, die Studien diagnostisch sinnvoll zu gruppieren. In der überwiegenden Mehrheit der Studien sind Patienten mit leichten bis mittelschweren Depressionen eingeschlossen, nur in einer der Studien erfolgte explizit eine Beschränkung auf Patienten mit schwerer Depression (486). Die Behandlungsdauer lag zwischen 4 und 8 Wochen.

2.3.2.3 Studienqualität

Die Mehrzahl der eingeschlossenen Studien war von guter methodischer Qualität. 19 Studien erreichten 4 oder 5 von 5 möglichen Punkten bei der Beurteilung mit der Skala von Jadad et al. Mit Ausnahme von 3 älteren Studien waren alle doppelblind. Die Verblindung der Randomisation

(concealment) war bei vielen Studien beschrieben und adäquat. In allen Fällen erfolgte die Umsetzung durch fortlaufend nummerierte neutrale Medikationscontainer. Bei den älteren Studien waren jedoch aus psychiatrischer Sicht in vielen Fällen Vorbehalte anzumelden. Die Studienpopulationen bzw. diagnostischen Kriterien waren häufig nur ungenau beschrieben, Vergleichsmedikationen niedrig dosiert und bzgl. der Beurteilungsinstrumente stellte sich die Frage einer angemessenen Anwendung. So fällt z.B. bei zwei Studien auf, dass hohe HAMD-Baselinewerte und gleichzeitig eine Diagnose leichte bis mittelschwere Depression berichtet werden (399). Ohne dass dies in den Jadadscores deutlich wird, erscheint insgesamt die Qualität in neueren Studien besser.

2.3.2.4 Studienergebnisse

Hypericumextrakte vs. Placebo

Insgesamt 22 Studien mit 2178 Patienten präsentierten Daten zum Therapieerfolg (Response, siehe Abbildung 8). Die gepoolte Responderratio lag bei 1,89 (95%-Konfidenzintervall 1,52 –2,36, $p < 0,00001$), d.h. die Hypericumextrakte erwiesen sich im Mittel Placebo deutlich überlegen. Der Test auf Heterogenität war hochsignifikant, d.h. die Studienergebnisse unterschieden sich überzufällig. Die Heterogenität blieb auch erhalten, wenn in Subgruppen von Präparaten getestet wurde. Die Responderratios waren hier für den Extrakt LI 160 (in zwei Versionen) 1,87 (1,35-2,60), für das Präparat Psychotonin (ebenfalls in zwei Versionen) 2,98 (1,13-7,84) und für die übrigen getesteten Extrakte 1,67 (1,19-2,36), d.h. für alle drei Subgruppen ergab sich wiederum eine signifikante Überlegenheit gegenüber Placebo. Ordnet man die Studien entsprechend dem Publikationsjahr, fällt auf, dass in neueren Studien der Effekt über Placebo hinaus zwar immer noch hochsignifikant, jedoch deutlich geringer ausgeprägt ist als in den älteren Studien (Abbildung 9). Abbildung 10 legt nahe, dass zwischen Schweregrad der Symptomatik vor Behandlung und der Effektivität von Hypericumextrakten ein Zusammenhang besteht: In den Studien an Patienten mit leichten Depressionen (HAMD-Score 12-18) ist die gepoolte Responderratio 2,70 (1,56-4,68), bei Patienten mit HAMD-Scores zwischen 19 und 21 2,02 (1,39-2,94) und bei Patienten mit HAMD-Scores über 22 1,22 (1,05-1,41). Allerdings sind neuere Studien häufiger an Patienten mit vergleichsweise schwerer Symptomatik durchgeführt worden. In 2 von 3 Subgruppen ist der Test auf Heterogenität weiterhin signifikant. Auch in den Analysen zu HAMD-Scorewerten sowie zu den Ergebnissen bei der Selbstbeurteilungsskala D-S (Depressionsskala, von Zerssen) erwiesen sich die Hypericumextrakte gegenüber Placebo deutlich überlegen (ohne Abbildung).

Comparison: 01 Hypericum mono-preparations vs. placebo A. Dichotomous measures

Outcome: 01 Responder

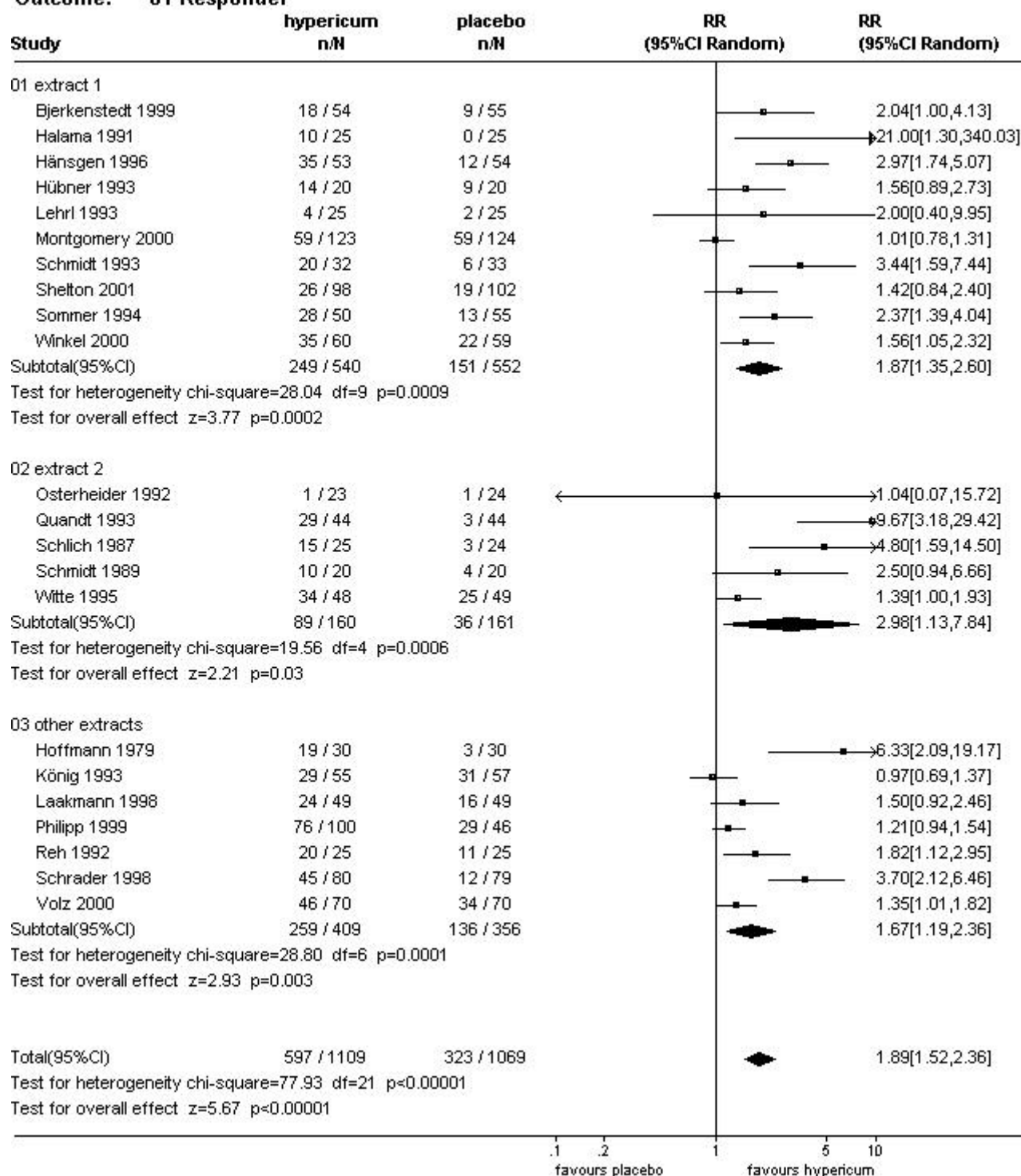


Abbildung 8:

Responderratios in Studien Hypericum vs. Placebo – geordnet nach Extrakten

(Abbildung aus Cochrane Collaboration Review Manager Software, Version 4.1)

n = Anzahl Responder, N = Anzahl Patienten in der jeweiligen Gruppe, RR = Responderratio, 95%CI random = 95% Konfidenzintervall

Extract 1 = LI 160 (Jarsin® oder Jarsin 300®), extract 2 = Psychotonin M® oder Psychotonin forte®

Comparison: 01 Hypericum mono-preparations vs. placebo A. Dichotomous measures

Outcome: 06 Responder - order by year with subgroups

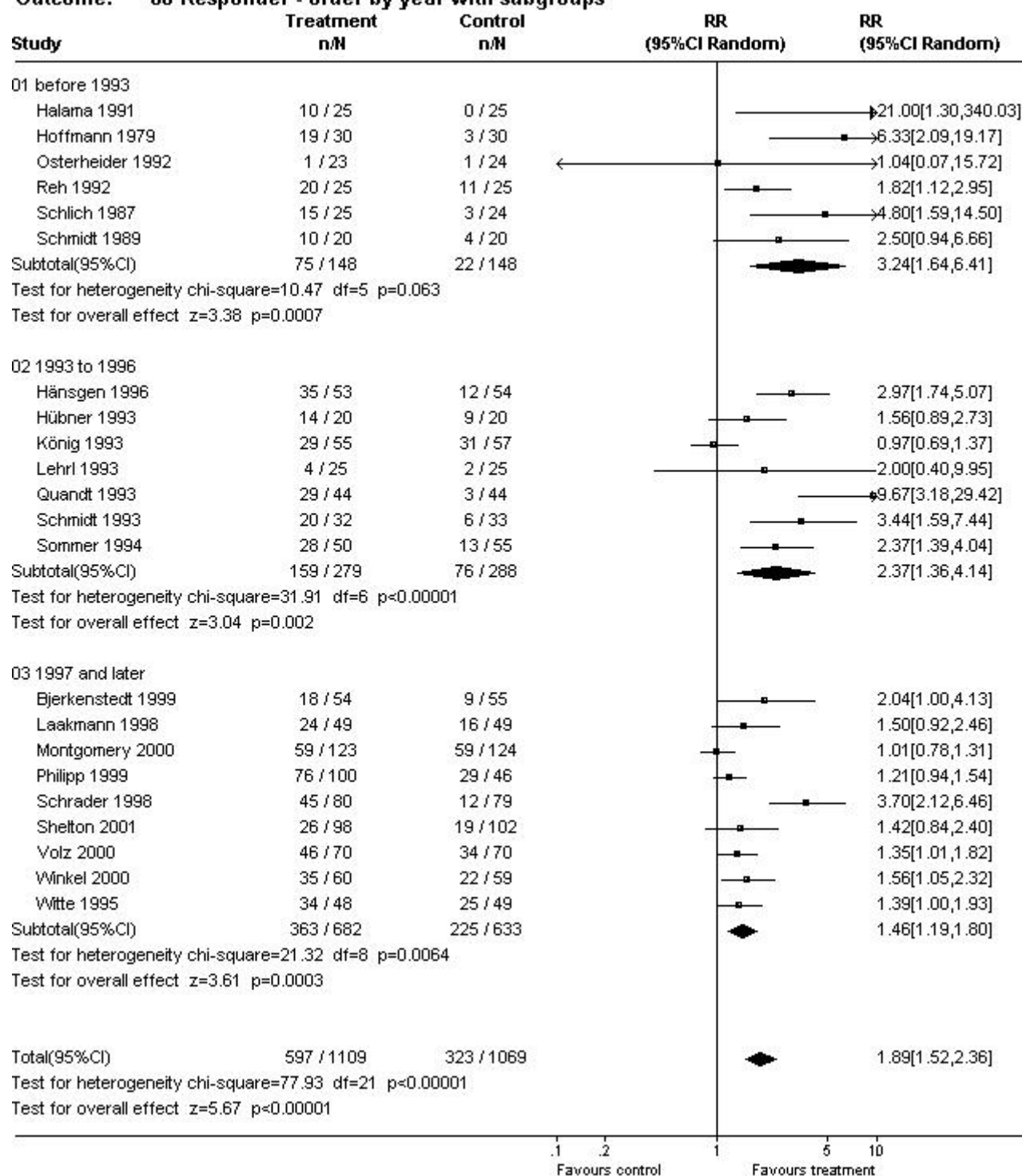


Abbildung 9:

Responderratios in Studien Hypericum vs. Placebo – geordnet nach Publikationsjahr

(Abbildung aus Cochrane Collaboration Review Manager Software, Version 4.1)

n = Anzahl Responder, N = Anzahl Patienten in der jeweiligen Gruppe, RR = Responderratio, 95%CI random = 95% Konfidenzintervall

Comparison: 01 Hypericum mono-preparations vs. placebo A. Dichotomous measures

Outcome: 08 Responder - order by HAMD baseline scores with subgroups

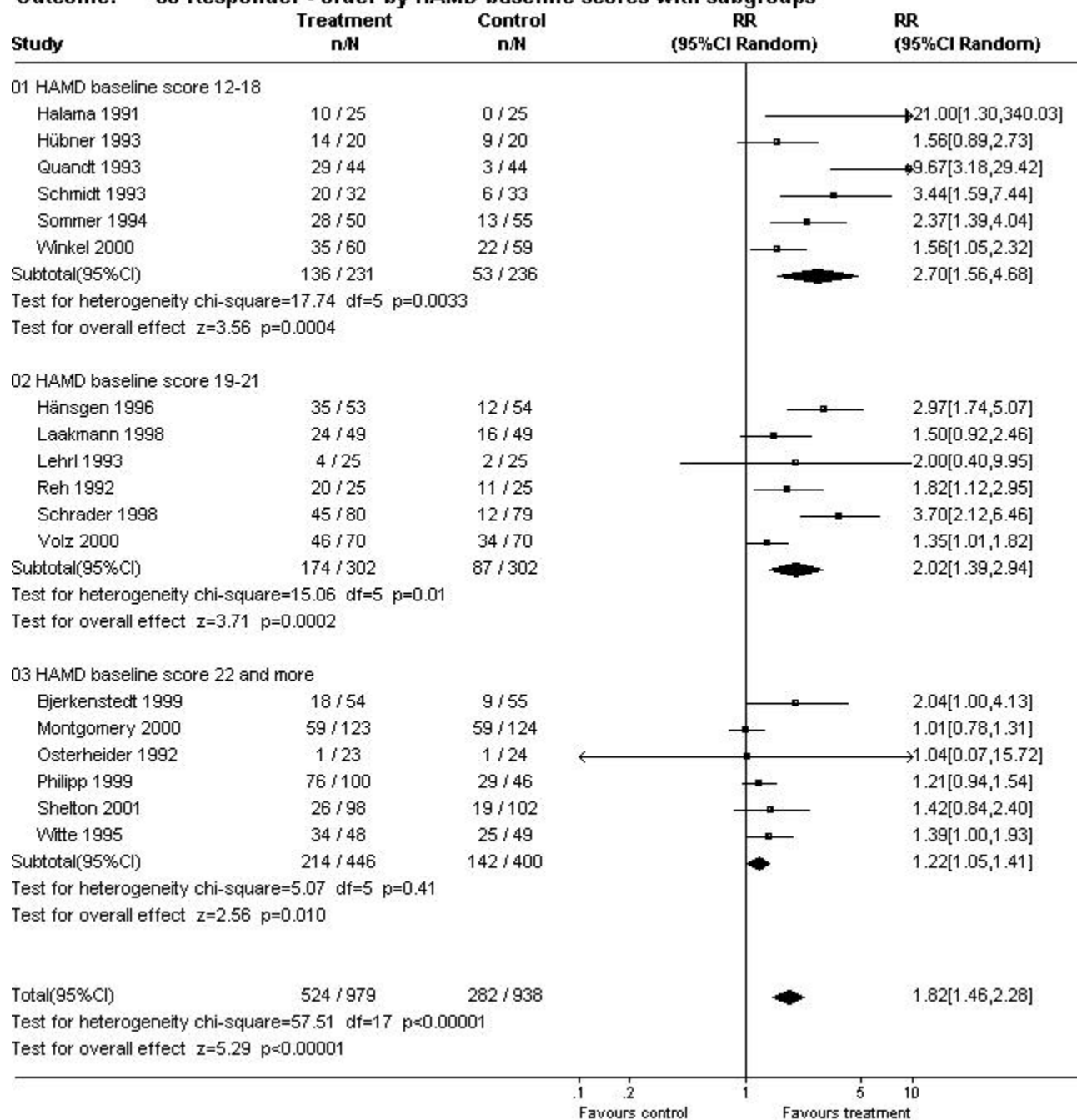


Abbildung 10:

Responderratio in Studien Hypericum vs. Placebo – geordnet nach Schweregrad (HAMD-Scorewerte) vor Behandlung
(Abbildung aus Cochrane Collaboration Review Manager Software, Version 4.1)

n = Anzahl Responder, N = Anzahl Patienten in der jeweiligen Gruppe, RR = Responderratio, 95%CI random = 95% Konfidenzintervall

Hypericumextrakte vs. andere Medikamente (Abbildung 11, Tabelle A12 im Anhang)

Für alle Studien konnten Responderratios berechnet werden. Die gepoolte Responderratio war 1,09 (95%-Konfidenzintervall 0,99-1,20), d.h. Hypericumextrakte waren in den einbezogenen Studien mindestens gleich effektiv wie die geprüften Antidepressiva. Die Subgruppenvergleiche zwischen dem Extrakt LI 160 und anderen Extrakten mit älteren Antidepressiva und selektiven

Serotonin-Wiederaufnahmehemmern erbrachten keine signifikanten Unterschiede. Nur in einer Studie (403) war der geprüfte Hypericumextrakt 20 mg Fluoxetin signifikant überlegen. Bei der Auswertung der HAMD-Scorewerte zeigte sich dieser Unterschied jedoch nicht. Klare Auswirkungen der Dosierung der Vergleichsantidepressiva oder des Schweregrades der Depression waren nicht zu erkennen.

Comparison: 05 Hypericum mono-preparations vs. standard antidepressants. A. Dichotomous measures
Outcome: 01 Responder

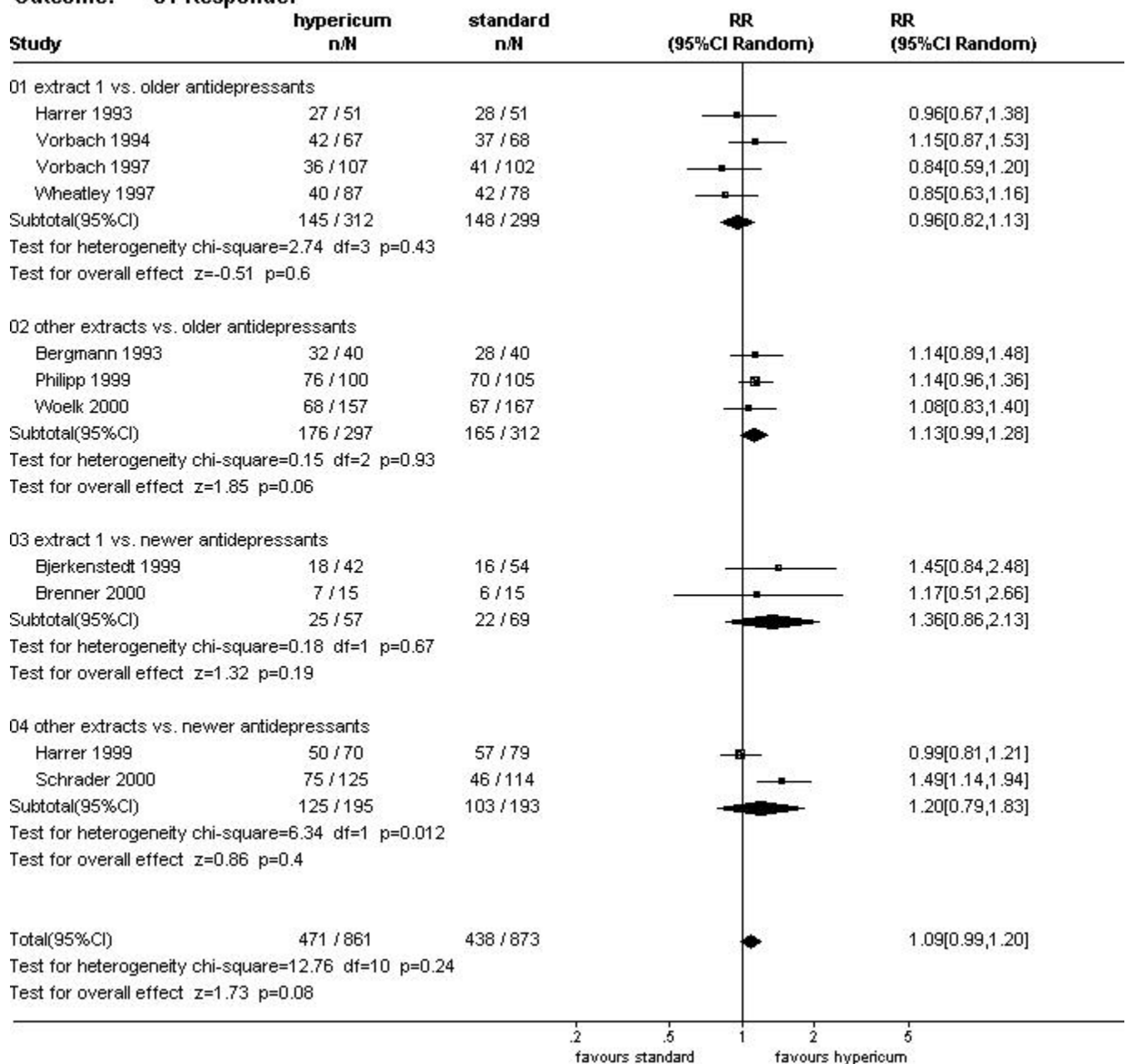


Abbildung 11:

Responderratios in Studien Hypericum vs. andere Medikamente

(Abbildung aus Cochrane Collaboration Review Manager Software, Version 4.1)

n = Anzahl Responder, N = Anzahl Patienten in der jeweiligen Gruppe, RR = Responderratio, 95%CI random = 95% Konfidenzintervall

Extract 1 = LI 160 (Jarsin® oder Jarsin 300®), extract 2 = Psychotonin M® oder Psychotonin forte®

Im Vergleich zu älteren Antidepressiva traten unter Hypericumextrakten bei deutlich weniger Patienten Nebenwirkungen auf (siehe Abbildung 12). Bei den zwei Vergleichen mit selektiven Serotonin-Wiederaufnahmehemmern war der Unterschied geringer ausgeprägt, in der gepoolten

Comparison: 09 Side effects - Hypericum mono-preparations vs. standard antidepressants
Outcome: 03 Number of patients reporting side effects

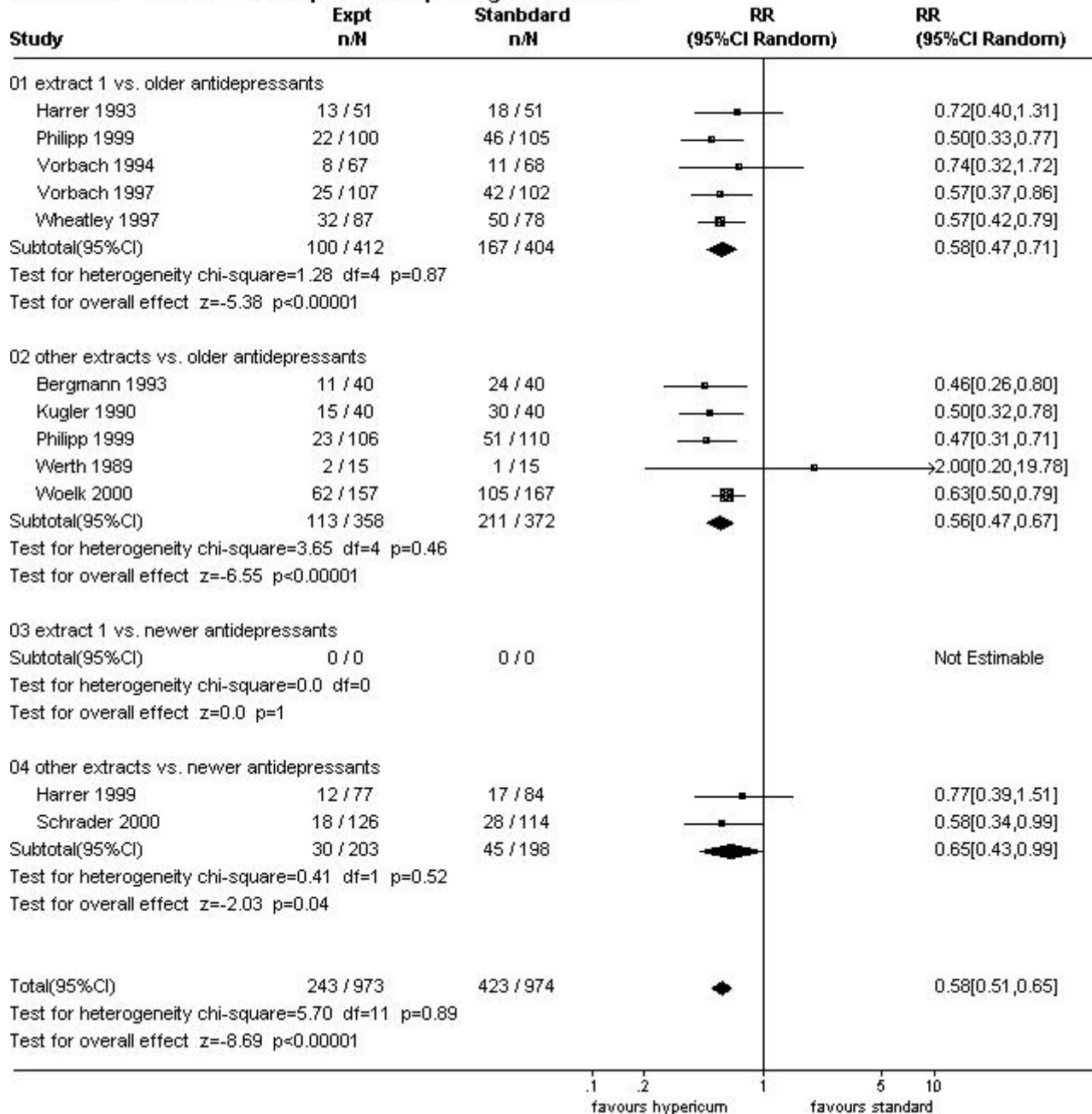


Abbildung 12:

Anzahl der Patienten mit Nebenwirkungen (relatives Risiko) in Studien Hypericum vs. andere Medikamente

(Abbildung aus Cochrane Collaboration Review Manager Software, Version 4.1)

n = Anzahl Responder, N = Anzahl Patienten in der jeweiligen Gruppe, RR = Responderratio, 95%CI random = 95% Konfidenzintervall
 Extract 1 = LI 160 (Jarsin® oder Jarsin 300®), extract 2 = Psychotonin M® oder Psychotonin forte®

Analyse jedoch immer noch signifikant zugunsten der Hypericumextrakte. Auch bezüglich der Anzahl von Studienabbrüchern (relatives Risiko 0,73, 95%-Konfidenzintervall 0,56-0,94) und der Anzahl von Studienabbrüchern wegen Nebenwirkungen (0,41; 0,26-0,66) schnitten die Hypericumextrakte signifikant besser ab.

Auffallend ist, dass alle 3 nicht in Deutschland durchgeführten placebokontrollierten Studien (338;407;467) keine signifikante Überlegenheit der geprüften Hypericumextrakte finden konnten. Dies könnte zufallsbedingt sein, könnte aber auch dafür sprechen, dass in Deutschland andere Patienten in die Studien einbezogen werden.

2.4 Diskussion

2.4.1 Migräneprophylaxe mit Propranolol, Metoprolol und Flunarizin

Die Wirksamkeit von Propranolol, Metoprolol und Flunarizin ist trotz der häufig unbefriedigenden Qualität der vorliegenden Studien im Vergleich zu Placebo gut belegt. Eindeutige Unterschiede zwischen diesen drei Substanzen sowie im Vergleich zu anderen Calciumantagonisten, β -Blockern sowie verschiedenen weiteren Substanzen lassen sich nur in Ausnahmefällen (z.B. der Überlegenheit von Metoprolol im Vergleich zur Acetylsalicylsäure (89;154)) erkennen. In den allermeisten Studien ist die Fallzahl jedoch so gering, dass aufgrund des Fehlens eines signifikanten Unterschiedes keineswegs auf Äquivalenz geschlossen werden kann. Aufgrund der unterschiedlichen Vergleichssubstanzen und Dosierungen, der häufig unbefriedigenden Studienqualität und der heterogenen oder unzureichenden Ergebnispräsentation erschien die Durchführung einer quantitativen Meta-Analyse obsolet. Vergleichende Aussagen über die Wirksamkeit einzelner Substanzen lassen sich daher schwer ableiten. Die Einschätzung von Propranolol, Metoprolol und Flunarizin als Migräneprophylaktika der ersten Wahl (21;87) lässt sich – wenn man sich ausschließlich auf die Evidenz aus randomisierten Studien beschränkt - daher eher durch die grundsätzlich belegte Wirksamkeit gegenüber Placebo als mit einer bewiesenen Überlegenheit im Vergleich zu anderen Substanzen begründen. Diese Schlußfolgerungen sind im Einklang mit denen der im Auftrag der Agency of Health Care Policy and Research durchgeführten Bewertung (62).

Aus methodischer Sicht besonders auffallend sind bei den eingeschlossenen Studien (1) die mangelhafte Beschreibung von Studienabbrüchen und deren Berücksichtigung in der Analyse sowie (2) die enorme Heterogenität und häufig mangelnde Detailgenauigkeit der Ergebnispräsentation.

Zu (1): In Studien zur Wirksamkeit einer medikamentösen Migräneprophylaxe kann es aufgrund der vergleichsweise langen Beobachtungsdauer, von Nebenwirkungen, wegen Nichtansprechens auf die Therapie, mangelnder Compliance der Patienten bzgl. Medikationseinnahme und Studienbedingungen (wie z.B. regelmäßige Arztbesuche und sorgfältige Führung des Kopfschmerztagebuchs) häufig dazu kommen, dass ein relevanter Anteil der Patienten die Studie vorzeitig abbricht oder ausgeschlossen wird. Solche Abbrüche und Ausschlüsse sind detailliert zu dokumentieren; diesbezügliche Empfehlungen wurden mehrfach publiziert (34;333). Leider fehlen z.T. auch in neueren Studien noch immer solche Angaben. In der Mehrheit der Studien ist aber darüberhinaus die Abbrecherrate so hoch (in der Größenordnung zwischen 10 und 50%), dass die Ergebnisse dadurch erheblich beeinflusst werden können, d.h. dass z.B. in placebokontrollierten Untersuchungen die Überlegenheit des Prüfpräparates z.T. überschätzt worden sein könnte. Analysen nach dem intent-to-treat Prinzip (514) wurden nur in wenigen Studien durchgeführt. Wenn solche Analysen durchgeführt wurden, wurden für fehlende Werte meist die zuletzt verfügbaren eingesetzt (last value carried forward). Bricht z.B. ein Patient nach 2 Monaten wegen Nebenwirkungen die Studie ab, werden bei dieser Vorgehensweise seine Angaben zur Attackenhäufigkeit in Monat 2 auch zur Bewertung von Monat 4 herangezogen. Bei Erkrankungen, bei denen nicht davon ausgegangen werden kann, dass sich die Ergebnisse bei längerer Beobachtungszeit verbessern, erscheint die Anwendung des last value carried forward Prinzips zumindest kritisierbar. Eine konservativere Einschätzung ergäbe sich vermutlich, wenn die fehlenden Werte durch Baselinewerte ersetzt würden. In zukünftigen Studien sollte unbedingt auf eine bessere Berichterstattung bzgl. Abbrechen und Ausschlüssen geachtet werden. Außerdem sollten sowohl per-protocol wie auch intent-to-treat Analysen präsentiert werden, wobei aus der Sicht des Autors bei den letzteren nicht nach dem last value carried forward Prinzip verfahren werden sollte.

Zu (2): Die Heterogenität in der Verlaufsbeurteilung und Ergebnisdarstellung macht eine systematische Bewertung der Ergebnisse fast unmöglich. Als Reviewer steht man vor einem „Flickenteppich“ von Resultaten, der für den Leser kaum transparent aufarbeitbar ist. Die Probleme sind einerseits in der Komplexität des Krankheitsbildes und den daraus folgenden Problemen bei der Quantifizierung der Symptomatik begründet, andererseits jedoch auch durch unzureichende Präsentation. Die Empfehlungen der International Headache Society für klinische Studien zur medikamentösen Behandlung der Migräne (198) geben einen definierten Katalog von Beurteilungskriterien, jedoch keine Präsentationsformate für die Publikation vor. Für und gegen die einzelnen Parameter und deren Darstellungsweisen gibt es jeweils eine Reihe von Argumenten. Dennoch erscheint es sinnvoll, bestimmte basale deskriptiv-statistische Angaben in Migräneprophylaxestudien standardmäßig zu präsentieren: 1) Jeweils ein zentraler Schätzer (Mittelwert oder Median) und ein Verteilungsmaß (Standardabweichung oder Quartile, nicht Standardfehler oder Konfidenzintervalle) für die Attackenzahl und/oder die Anzahl der Tage mit

Kopfschmerzen (dieser Parameter ist in Multicenterstudien zuverlässiger zu erheben, kann aber aufgrund von begleitenden Nicht-Migränekopfschmerzen Probleme bereiten) und für die Anzahl der Tage, an denen Medikamente zur Attackenkupierung eingenommen wurden; 2) Anzahl der Patienten, bei denen im Vergleich zur Baselinephase die Attackenhäufigkeit und/oder die Anzahl der Kopfschmerztag um 50% oder mehr abgenommen hat. Vier-Wochenperioden sind häufige Beurteilungszeiträume in Migräneprophylaxestudien. Daher sollten in jedem Fall Angaben für tatsächlich beobachtete Werte (nicht ausschließlich Abnahmen gegenüber Baseline in Prozent oder absolut) in den 4 Wochen vor Behandlung und in den letzten 4 Behandlungswochen gemacht werden. Weitere Zeitfenster sind selbstverständlich wünschenswert. Für die Berechnung von Effektmaßen für Attackenhäufigkeit und Kopfschmerztag in Meta-Analysen ist die Angabe von Mittelwerten und Standardabweichungen sowie die Anzahl der zugrundeliegenden Beobachtungen notwendig. Da die Daten häufig nicht normalverteilt sind, geben viele Autoren - grundsätzlich richtigerweise - den Median (und selten zusätzlich Quartile) an. In solchen Fällen wäre die zusätzliche Angabe von Mittelwerten und Standardabweichungen dennoch wünschenswert. Angaben zur Kopfschmerzdauer erscheinen wegen der fraglichen Verlässlichkeit problematisch. Für den Patienten ist es häufig schwer einzuschätzen, wann die Kopfschmerzen begannen und wann sie aufhörten. Schwierig sind auch Angaben zur Attackenintensität. Einerseits kann die Intensität der einzelnen Attacken gemittelt werden, auf der anderen Seite können die Intensitäten in einem Score addiert werden.

Zusammenfassend muß festgestellt werden, dass die klinischen Studien zur Migräneprophylaxe mit Propranolol, Metoprolol und Flunarizin die Wirksamkeit dieser Substanzen grundsätzlich belegen, dennoch aber viele Fragen offen lassen. Langzeitstudien sind selten oder aufgrund hoher Abbrecherraten kaum interpretierbar. Vergleiche verschiedener Medikamente haben meist keine ausreichende Power, um eine Äquivalenz abzusichern. Schließlich scheinen Autoren von Publikationen international gängige Empfehlungen zur Veröffentlichung randomisierter Studien (34;333) weitgehend zu ignorieren. Das Ausmaß der prophylaktischen Wirksamkeit der untersuchten Medikamente ist außerdem im Verhältnis zu den nicht unbeträchtlichen Nebenwirkungen nicht wirklich überzeugend, sodass die Suche nach anderen effektiven und nebenwirkungsarmen Behandlungsstrategien fortgesetzt werden sollte.

2.4.2 Behandlung von Depressionen mit Hypericumextrakten

Die Überlegenheit von Hypericumextrakten bei der Behandlung von leichten bis mittelschweren Depressionen im Vergleich zu Placebo ist durch zum Teil qualitativ hochwertige Studien gut belegt. Bei schweren Depressionen scheint die Wirksamkeit, zumindest bei den geprüften Dosierungen, weniger deutlich ausgeprägt zu sein. Außerdem sprechen die Daten dafür, dass das Ausmaß des Effektes in älteren Studien deutlich überschätzt wurde. Die vorhandene Evidenz spricht auch dafür, dass Hypericumextrakte bei leichten bis mittelschweren Depressionen ähnlich wirksam sind wie andere Antidepressiva, gleichzeitig jedoch weniger Nebenwirkungen haben. Aufgrund der häufig unzureichenden Dosierungen der Vergleichspräparate in den älteren Studien und der Vielzahl von Hypericumextrakten und konventionellen Antidepressiva, sind hier jedoch weitere Studien notwendig. Diese Schlussfolgerungen sind weitgehend in Übereinstimmung mit denen anderer systematischer Übersichtsarbeiten zum Thema (siehe Kapitel 4).

Die Interpretation der placebokontrollierten Studien wird durch die erhebliche Heterogenität der berechneten Effektgrößen erschwert. Die statistischen Tests auf Heterogenität werden allgemein als wenig sensitiv angesehen (461). Umso mehr müssen die hochsignifikanten Ergebnisse diesbezüglich beunruhigen. Ein plausibler Grund für die Heterogenität könnten theoretisch die unterschiedlichen Extrakte sein. Da jedoch auch innerhalb der Subgruppenanalysen zu einem Extrakttyp die Heterogenität nur unwesentlich abnimmt, kann dies nicht die einzige Erklärung sein. Aus Sicht des Autors erscheinen vor allem Unterschiede bzgl. Studienqualität und der Studienpopulationen plausible Gründe für die variierenden Effektgrößen. Die Qualitätsunterschiede sind jedoch mit gängigen methodischen Qualitätsskalen kaum zu objektivieren, da auch einige der subjektiv als problematisch eingestuften Studien hohe Scores erreichten. Eine weitergehende Untersuchung der Ursachen für die beobachtete Heterogenität durch Metaregressionstechniken, bei denen mehrere potentielle Einflussfaktoren gleichzeitig berücksichtigt werden können (414), wird für die Zukunft angestrebt.

Hypericumextrakte sind in jüngster Vergangenheit aufgrund von Interaktionen mit anderen Arzneimitteln in die Diskussion geraten (113). Zahlreiche Fallberichte zeigen, dass die Plasmakonzentration verschiedener Wirkstoffe wie z.B. Cyclosporin (391), Indinavir (369) oder Warfarin (527) bei gleichzeitiger Hypericumgabe abnimmt. Als Wirkungsmechanismus wird vor allem eine Induktion von Cytochrom P-450-Enzymen diskutiert (113). Dies verdeutlicht, dass Hypericumextrakte trotz ihrer grundsätzlich sehr guten Verträglichkeit nur unter ärztlicher Aufsicht eingenommen werden sollten.

Mit Spannung werden derzeit die Ergebnisse einer von den National Institutes of Health (NIH) gesponsorten Studie erwartet, in der die Wirksamkeit eines Hypericumextrakts, eines selektiven

Serotonin-Wiederaufnahmehemmern und von Placebo bei mehr als 300 Patienten mit majorer Depression verglichen wird. Dieser Studie wird besonders große Bedeutung beigemessen, a) da sie vermutlich die erste nicht von einer Pharmafirma gesponsorte Studie zu einem Hypericumpräparat ist, b) weil sie mit besonders hohem finanziellen und methodischem Aufwand durchgeführt wird und c) da die erste placebokontrollierte amerikanische Studie von Shelton et al. (407) ein negatives Ergebnis erbracht hat. Diese letzte, methodisch sehr hochwertige Studie wurde allerdings kontrovers diskutiert, da sie 1) Patienten einschloss, die im Mittel seit mehr als 2 Jahren an der aktuellen depressiven Episode litten und 2) von einer Pharmafirma gesponsort wurde, die kein Hypericumpräparat herstellt, dafür jedoch einen selektiven Serotonin-Wiederaufnahmehemmer, der aber nicht in die Prüfung miteingebracht wurde. Sollte die NIH-Studie keine Überlegenheit im Vergleich zu Placebo bei gleichzeitiger Unterlegenheit zum synthetischen Antidepressivum erbringen, ist damit zu rechnen, dass die bisher vorliegenden, hauptsächlich aus Deutschland stammenden, positiven Ergebnisse international in Frage gestellt werden. Allerdings ist auch hier nochmals zu berücksichtigen, dass in der NIH-Studie nur Patienten mit majorer Depression eingeschlossen wurden und die vorliegende Meta-Analyse darauf hinweist, dass bei diesen Patienten der Effekt geringer ausgeprägt ist als bei Patienten mit leichteren Depressionen.

Ob mit höheren Dosierungen bei schwereren Depressionen bessere Ergebnisse erzielt werden könnten, ist eine Fragestellung, die angesichts der guten Verträglichkeit von Hypericumextrakten untersucht werden sollte. Bisher wurde nur in einer einzigen Studie (486) eine vergleichsweise hohe Extraktdosis untersucht.

Trotz dieser Diskussionen erscheint die Gabe qualitativ hochwertiger und in klinischen Studien geprüfter Hypericumextrakte zur Behandlung leichter bis mittelschwerer Depressionen gerechtfertigt. Weitere sorgfältige Studien müssen zeigen, inwieweit Hypericumextrakte bei majoren bzw. bei schwereren Depressionen vertretbar sind.

3 Empirisch-methodische Untersuchungen zur Studienqualität und deren Einfluss auf das Studienergebnis

3.1 Hintergrund

Die Beurteilung der Qualität der Primärstudien wird allgemein als ein entscheidender Bestandteil einer hochwertigen systematischen Übersichtsarbeit angesehen (468). Bereits bzgl. der Definition von Qualität gibt es jedoch erhebliche Schwierigkeiten. In Reviews von Therapiestudien wird Qualität häufig als methodische Qualität oder interne Validität interpretiert. Eine entsprechende Definition ist dann z.B. „Qualität ist die Wahrscheinlichkeit, mit der eine Studie unverzerrte Ergebnisse erbringt“ (206). Es herrscht weitgehend Konsens, dass die diesbezüglich relevantesten Kriterien Randomisation, Verblindung und Vermeidung von Verzerrungen bei der Auswertung sind (225).

Eine methodisch saubere Studie muss jedoch nicht in jedem Fall eine gute Studie sein. Sie kann z.B. klinisch irrelevant (d.h. für die Entscheidungsfindung unwichtig) oder nicht extern valide sein (d.h. die Ergebnisse sind nicht auf die Mehrheit der Patienten außerhalb der Stichprobe übertragbar). Auch hier gibt es zwar grundsätzlich relevante Kriterien (z.B. erhöht eine zufällige Stichprobenziehung die Generalisierbarkeit), jedoch ist deren Operationalisierung in hohem Maße von den spezifischen Umständen abhängig. D.h. es ist kaum möglich, unabhängig von Interventionen und Indikationen valide Kriterien zur Bewertung der externen Validität und klinischen Relevanz zu definieren.

Zur Bewertung der internen Validität oder methodischen Qualität liegt eine ganze Reihe von Skalen und Checklisten vor (siehe (329) für eine nicht mehr ganz aktuelle Übersicht). Die meisten dieser Skalen definieren jedoch nicht explizit, was sie messen wollen, ihre Entwicklung war in der Regel nicht systematisch und nur eine ist umfangreich validiert worden (206). Diese Skala von Jadad et al. wurde auch vom Autor in allen in dieser Schrift enthaltenen Arbeiten – zum Teil neben anderen Skalen - verwendet. Die Tatsache der Validierung darf aber nicht darüber hinwegtäuschen, dass es sich bei dieser Skala, ebenso wie bei den meisten anderen Skalen, nur um ein äußerst grobes Werkzeug handelt (siehe 3.4.). Bei vielen der Skalen wird ein Summenscore als primärer Qualitätsindikator berechnet. Die Sinnhaftigkeit solcher Qualitätsscores ist jedoch in hohem Maße umstritten (152;225;352).

Die erste empirisch-methodische Untersuchung in diesem Kapitel widmet sich der Bewertung der methodischen Qualität randomisierter klinischer Studien in den Bereichen Akupunktur, Phytotherapie und Homöopathie. Die Bewertung stützt sich sowohl auf Scores wie auch auf Einzelkriterien. Ziel war zu untersuchen, 1) wie die methodische Qualität dieser Studien

einzuschätzen ist und wo besondere Schwachstellen liegen, 2) ob es zwischen den drei Therapieformen relevante Unterschiede gibt und 3) ob die methodische Qualität mit Ort, Sprache und Zeitpunkt der Publikation sowie mit der Fallzahl zusammenhängt.

Ein logischer nächster Schritt ist zu prüfen, ob methodisch schlechtere Studien zu anderen Ergebnissen kommen als methodisch bessere Studien. Diesem Thema widmet sich die zweite Untersuchung in diesem Kapitel am Beispiel der placebokontrollierten Studien zur Homöopathie. In der Literatur finden sich mehrere Untersuchungen, die bei randomisierten Studien zu „schulmedizinischen“ Interventionen deutliche Hinweise dafür gefunden haben, dass qualitativ hochwertigere Studien zu weniger positiven Ergebnissen kommen (wichtige Beispiele (331;406)). Hierbei stehen unterschiedliche Methoden zur Verfügung. Ein Ansatz ist, Studien zu vergleichen, die bei der Qualitätsbewertung mit einer Skala über oder unter einer bestimmten Mindestpunktzahl („cut-off-point“) liegen (verwendet z.B. in (331)). Bei einer größeren Anzahl von Primärstudien ist es grundsätzlich möglich, die Studien nach und nach entsprechend ihrem Qualitätsscore in die Metaanalyse einzuschliessen (d.h. zuerst nur die Studien mit den höchsten Scores und dann nach und nach Studien mit immer niedrigeren Scores). Dieses Vorgehen wird als kumulative Metaanalyse bezeichnet. Schließlich besteht die Möglichkeit den Einfluss einzelner Qualitätsindikatoren auf das Ergebnis – nach Möglichkeit unter Berücksichtigung des Einflusses der anderen Qualitätsindikatoren – zu untersuchen (414). Alle drei Vorgehensweisen wurden bei den placebokontrollierten Studien zur Homöopathie angewendet und verglichen.

Die dritte Untersuchung in diesem Kapitel erweitert die Problematik von Qualitätsbewertung und Einfluss der Qualität auf das Studienergebnis auf nichtrandomisierte Studien und Aspekte der externen Validität und Interpretierbarkeit. Die überwiegende Mehrheit der systematischen Übersichtsarbeiten zu präventiven und therapeutischen Interventionen ist auf randomisierte Studien beschränkt (102). Eine kleine Minderheit bezieht auch nichtrandomisierte, kontrollierte Studien mit ein (199). Unkontrollierte Studien werden fast nie einbezogen. Zuverlässige Aussagen zur Wirksamkeit sind aus nichtrandomisierten kontrollierten Studien nur mit Vorsicht, aus unkontrollierten Studien nur in Ausnahmefällen abzuleiten. Randomisierte Studien haben auf der anderen Seite aber auch klare Nachteile. In randomisierten Studien kann z.B. oft nur ein Bruchteil der geeigneten Patienten einbezogen werden, da die Mehrheit die Zustimmung verweigert (317). Dadurch sind zum einen große Fallzahlen schwer zu verwirklichen, zum anderen ist die Repräsentativität und damit die externe Validität fragwürdig. Vor allem wenn ein Teil der Patienten zu Nichtbehandlung oder Placebo randomisiert wird, sind Langzeitstudien schwierig, da viele Patienten die Studie abbrechen und eine Behandlung ihrer Wahl beginnen. *Sorgfältig geplante und durchgeführte* nichtrandomisierte Studien mit großen und repräsentativen Stichproben oder langer Nachbeobachtungsphase (Follow-up) können daher äußerst wichtige, *ergänzende* Informationen erbringen, z.B. bzgl. prognostisch relevanter Faktoren, seltenen Nebenwirkungen oder

Komplikationen, der Übertragbarkeit von Ergebnissen randomisierter Studien auf die Routinepraxis und zu Langzeitverläufen. Der Einbezug von nichtrandomisierten – und dabei vor allem nichtkontrollierten – Studien stellt den Reviewer aber vor große Herausforderungen bzgl. Literatursuche (keine gleichzeitig sensitiven und spezifischen Suchstrategien für die großen Datenbanken verfügbar), Bewertung (keine Instrumente zur Qualitätsbewertung über ein Spektrum von Studiendesigns) und Auswertung (einerseits Zwischengruppenunterschiede aus kontrollierten Studien, andererseits Innergruppenunterschiede aus allen Studien). Erfahrungen und empirisch-methodische Untersuchungen zu diesem Problemkreis liegen im Bereich der Medizin - anders als z.B. in der Psychologie oder den Sozialwissenschaften - kaum vor. Am Beispiel von klinischen Studien unterschiedlicher Designs zur Akupunktur bei chronischen Kopfschmerzen wurden folgende in diesem Zusammenhang relevante Fragen bearbeitet: 1. Unterscheiden sich randomisierte und nichtrandomisierte Studien zur Akupunktur bei chronischen Kopfschmerzen bzgl. Patienten, Interventionen, designübergreifenden Qualitätsaspekten und Therapieerfolg? 2. Erbringen nichtrandomisierte Studien relevante Zusatzinformationen bzgl. Langzeitverlauf, prognostischen Faktoren, Nebenwirkungen und Komplikationen sowie bzgl. Generalisierbarkeit? 3. Falls Unterschiede zwischen Therapieerfolgen in randomisierten und nichtrandomisierten Studien bestehen sollten, was sind die Gründe?

3.2 Methodik

3.2.1 Methodische Qualität randomisierter Studien in den Bereichen Akupunktur, Phytotherapie und Homöopathie

3.2.1.1 Studienselektion

Die in der präsentierten Analyse verwendeten Daten wurden ursprünglich im Rahmen von systematischen Übersichtsarbeiten zur Homöopathie (278), zu Johanniskrautpräparaten bei Depression (285;286), Echinaceapräparaten bei Erkältungskrankheiten (320), Akupunktur bei Asthma (280;290) und Akupunktur bei chronischen Kopfschmerzen erhoben (319;321). Die Literatursuche für diese Übersichtsarbeiten erfolgte durch Abfragen in Datenbanken (Medline, Embase, CISCOR, Phytodok, Datenbank des Zentrums für naturheilkundliche Forschung München), durch Prüfen von Bibliographien und Literaturverzeichnissen gefundener Arbeiten sowie durch Kontakte mit Experten und z.T. Herstellern. Für die vorliegende Analyse wurden nur kontrollierte Studien berücksichtigt, die folgende Kriterien erfüllten: 1) Die Behandlung der Studienteilnehmer diene präventiven oder therapeutischen Zwecken (also z.B. keine Studien an Gesunden mit ausschliesslicher Testung physiologischer Größen oder z.B. der Pharmakokinetik). 2) Die Zuteilung zu Prüf- und Kontrollgruppen erfolgte randomisiert oder quasirandomisiert (z.B.

durch Alternierung). Ausdrücklich als doppelblind bezeichnete Studien, bei denen die Art der Zuteilung nicht explizit beschrieben war, wurden eingeschlossen, wenn davon auszugehen war, dass die einschließende Person aufgrund der Blindbedingungen nicht wußte, welcher Gruppe der jeweilige Patient zugeteilt werden würde. 3) Die Prüfbehandlung wurde mit Placebo, keiner Behandlung oder einer anderen Behandlung verglichen (d.h. Studien, in denen z.B. ausschließlich zwei Dosen eines Präparates verglichen wurden, wurden nicht berücksichtigt). Nicht berücksichtigt wurden außerdem Studien, die in einem Abstract oder einer Publikation zwar erwähnt wurden, zu denen jedoch keinerlei Details verfügbar waren. Details zu den einzelnen Übersichtsarbeiten können den jeweiligen Publikationen entnommen werden.

Insgesamt 207 Studien (129 zur Homöopathie, 27 zu Johanniskrautpräparaten bei Depression, 15 zu Echinaceapräparaten bei Erkältungskrankheiten, 15 zu Akupunktur bei Asthma und 21 zu Akupunktur bei Kopfschmerzen) erfüllten die Einschlußkriterien. Für die vorliegende Analyse wurden folgende Angaben aus den für die einzelnen Übersichtsarbeiten erstellten Datenbanken in einer SPSS-Datei (SPSS Corp., Chicago, Illinois, USA) zusammengeführt: Art der Quelle (Artikel in Medline-gelisteter Zeitschrift, sonstiger Zeitschrift, andere Veröffentlichungen, unveröffentlichter Bericht), Publikationsjahr, Herkunftsland (Land der Institution, an der der Erstautor tätig war), Publikationssprache (englisch oder andere Sprache), Fallzahl, Art der Kontrollintervention sowie Daten zu Studienqualität.

3.2.1.2 Beurteilung der Studienqualität

Primäres Instrument zur Beurteilung der methodischen Qualität war die von Jadad et al. entwickelte und validierte Skala ((206) Jadadscore). Die Skala umfasst 3 Items zur Beschreibung von methodischen Schlüsselaspekten, die mit insgesamt 5 Punkten bewertet werden können (weitere Details siehe 2.2.1.4). Zusätzlich wurde eine vom Autor entwickelte Checkliste eingesetzt (Internal Validity Scale). Dieses Instrument, das unabhängig von und gleichzeitig mit dem Score von Jadad et al. erarbeitet worden war, wurde mitgeführt, da es aus der Sicht des Autors zwar weniger systematisch validiert, aber konzeptionell schlüssiger war und da es zusätzliche Komponenten enthielt. Die Internal Validity Scale dient der Abschätzung, ob ggf. beobachtete Unterschiede in einer kontrollierten klinischen Studie tatsächlich auf die unabhängige Variable zurückzuführen sind. Die erste Entwicklung der Skala erfolgte auf der Basis publizierter Skalen und Checklisten. Vorläufige Versionen der Skala wurden in Sets von 10 bis 30 Studien auf Praktikabilität, Reliabilität und „face validity“ geprüft und dann modifiziert. Die endgültige Version umfasst 6 Items, deren Einfluß auf das Studienergebnis sowohl einzeln bzw. in Untergruppen wie auch als Gesamtscore geprüft werden kann. Beurteilt werden die Methode der Zuteilung („treatment allocation“), die Verblindung der Zuteilung („allocation concealment“), die Vergleichbarkeit der Gruppen bei Studienbeginn („baseline comparability“), die Verblindung der Patienten („blinding of patients“), die Verblindung der Beurteiler („blinding of evaluators“) und der

Umgang mit Studienabbrüchen und –ausschlüssen („handling of drop-outs and withdrawals“). Für jedes Item kann maximal ein Punkt vergeben werden, also insgesamt sechs. Die Antwortmöglichkeiten sind vorgegeben und durch eine Bewertungsanleitung näher spezifiziert. Diese unterscheidet sich geringfügig bei medikamentösen und nichtmedikamentösen Studien. Bei den ersteren wird Item 5 („blinding of evaluators“) nur als erfüllt gewertet, wenn auch der Therapeut (der meist auch Beurteiler ist) verblindet ist. In der vorliegenden Präsentation wird jedoch primär nur das Item für „allocation concealment“ miteinbezogen.

Jede klinische Studie wurde von mindestens 2 Reviewern unabhängig mit Hilfe von vorgegebenen Formularen bearbeitet. Diese wurden vom Autor auf Nichtübereinstimmungen geprüft; relevante Diskrepanzen wurden diskutiert und im Konsens gelöst. Insgesamt wirkten 9 Beurteiler mit; der Autor beurteilte alle Studien. Die Übereinstimmung vor Diskussion (berechnet für die vier Reviewer, die mehr als 10 Studien beurteilten) war für beide Qualitätsskalen gut bis sehr gut (Intraclass-Korrelations-Koeffizienten zwischen 0,65 und 0,96).

3.2.1.3 Auswertung

Die Daten aus den 5 Übersichtsarbeiten wurden in einer SPSS-Matrix zusammengeführt. Zur Prüfung auf Unterschiede in Merkmalen und Qualität der Studien zu Homöopathie, Phytotherapie und Akupunktur wurde der Kruskal-Wallis- bzw. der Chi-Quadrat-Test verwendet, für Differenzen zwischen Studien zu Echinacea und Johanniskraut bzw. Akupunktur bei Asthma und Kopfschmerzen der Wilcoxon-Mann-Whitney-Test und der Chi-Quadrat-Test. Um zu überprüfen, ob die Werte für den Jadadscore mit der Publikationssprache (publiziert in Englisch oder nicht), Publikationsort (in einer Medline-gelisteten Zeitschrift oder nicht), Publikationszeitpunkt (vor oder ab 1990), oder mit der Studiengröße (Fallzahl von unter oder mindestens 100 Patienten) assoziiert sind, wurden unadjustierte und adjustierte (allgemeines lineares Regressionsmodell mithilfe der GLM-Prozedur in SPSS) Scores berechnet. Korrelationen zwischen einzelnen methodischen Qualitätskriterien und den obengenannten Einflussfaktoren wurden durch logistische Regression geprüft. Bei Tests über alle drei Therapien (Akupunktur, Phytotherapie und Homöopathie) wurde nur eine zufällige Stichprobe (gezogen mithilfe der entsprechenden Funktion in SPSS) der Homöopathiestudien miteinbezogen, um zu vermeiden, dass der überproportionale Anteil dieser Studien an der Gesamtstudiensammlung die Ergebnisse beeinflusst. P-Werte < 0,05 wurden als statistisch signifikant angesehen; Adjustierungen für multiples Testen erfolgten nicht.

3.2.2 Einfluss von Aspekten der methodischen Qualität auf das Studienergebnis am Beispiel der placebokontrollierten Studien zur Homöopathie

3.2.2.1 Studienselektion

Die Methoden der von uns durchgeführten Meta-Analyse der placebokontrollierten Studien zur Homöopathie sind an anderer Stelle ausführlich beschrieben (278). An dieser Stelle wird daher lediglich auf die Aspekte detailliert eingegangen, die den Zusammenhang zwischen Studienqualität und –ergebnis untersuchen.

Placebokontrollierte Studien zur Homöopathie (im Gegensatz zu 3.2.1. wo auch nicht-placebokontrollierte miteinbezogen wurden) wurden durch Suchen in Medline, Embase, komplementärmedizinischen Datenbanken, Kontakt mit Experten und Prüfung der Literaturverzeichnisse von klinischen Studien und Übersichtsartikeln identifiziert. Eingeschlossen wurden randomisierte und/oder doppelblinde Studien (d.h. wenn bei einer Studie eine klare Beschreibung der Zuteilungsmethode fehlte, wurde diese nur eingeschlossen, wenn die Studie explizit als doppelblind bezeichnet war und damit davon auszugehen war, dass der aufnehmende Arzt nicht wissen konnte, welche Behandlung der jeweils nächste Patient erhalten würde), in denen die Wirksamkeit einer homöopathischen Intervention (Verwendung potenziierter homöopathischer Arzneimittel, also keine Urtinkturen) im Vgl. zu Placebo bei prophylaktischer oder therapeutischer Behandlung beim Menschen untersucht wurde.

Insgesamt 119 Studien entsprachen diesen Kriterien, 30 präsentierten jedoch keine für eine Meta-Analyse verwertbaren Daten (d.h. ausreichende Angaben zu einem der folgenden Zielkriterien: vordefiniertes Hauptzielkriterium; wenn kein Hauptzielkriterium definiert war, Globalbeurteilung durch Patienten oder Ärzte; wenn zu keinem der genannten Parameter verwertbare Daten vorhanden waren, dann nach Möglichkeit eindeutig identifizierbares wichtigstes klinisches Kriterium). Alle Erhebungen erfolgten unabhängig durch zwei Reviewer mit Hilfe eines vorbereiteten Formulars.

3.2.2.2 Beurteilung der methodischen Qualität

Die methodische Qualität bzw. die interne Validität wurden durch die in 2.2.1.4. bzw. 3.2.1.2. beschriebenen Instrumente Jadadscore und Internal Validity Scale beurteilt. Bei der Internal Validity Scale wurde ein zusätzliches Item zur Angemessenheit der statistischen Analyse in den Studien mitgeführt, da für eine mit der Untersuchung gekoppelte Dissertation (71) auch alle vorhandenen kontrollierten Studien bewertet wurden, ohne dass deren Ergebnisse in einer Meta-Analyse zusammengefaßt wurde. Die Internal Validity Scale umfaßt in der folgenden Analyse

daher 7 Items. Wiederum beurteilten mindestens zwei Reviewer jede Studie mit beiden Instrumenten.

3.2.2.3 Auswertung

Für jede Studie wurde eine Odds Ratio (OR) mit den zugehörigen 95%-Konfidenzintervallen berechnet; ORs > 1 bedeuten ein Ergebnis zugunsten der homöopathischen Intervention, ORs < 1 ein Ergebnis zugunsten von Placebo. Für die vorliegende Auswertung wurden die Daten der Originalübersicht (278) mit Hilfe von Meta-Regressions-Methoden reanalysiert, die die Zwischen-Studien-Heterogenität berücksichtigen. In einem ersten Schritt wurde der natürliche Logarithmus der ORs gebildet. Dementsprechend ergibt sich ein Behandlungseffekt $y_i = \ln(OR)$ für jede Studie i mit $i = 1, \dots, n$. Folgendes Meta-Regressionsmodell wurde zugrundegelegt:

$$Y_i = x_i\beta + b_i + \varepsilon_i,$$

wobei x_i für einen Vektor bekannter Kovariablen für die Studie i steht, $b_i \sim N(0, \tau^2)$ für einen studienspezifischen Zufallseffekt auf der Basis der residualen Zwischen-Studien-Heterogenität und $\varepsilon_i \sim N(0, \sigma_i^2)$ mit der jeweiligen Varianz der einzelnen Studie σ_i^2 . Diese wurde anhand der berechneten Varianz s_i^2 der i -ten Studie berechnet. Wenn b_i und ε_i unabhängig sind, ist $\text{var}(y_i)$ gleich $\tau^2 + s_i^2$. Unter dieser Modellannahme ist der geschätzte lineare Vorhersagewert $x_i\beta$ der mittlere Effekt aller Studien mit dem Kovariatenvektor x_i und b_i die Abweichung der i -ten Studie von diesem Mittelwert. Die Parameter-Schätzer wurden mit Hilfe der restricted maximum likelihood (REML) Methode in der PROC MIXED Funktion in SAS 6.12 (SAS Inst., Cary, North Carolina, USA) bestimmt.

Zur Überprüfung der Hypothese, dass keine residuale Heterogenität zwischen den Studien mehr besteht, wurde ein likelihood ratio Test durchgeführt, in dem die Wahrscheinlichkeit dieses Modells mit einem Modell, in dem τ^2 geschätzt wird, verglichen wird. Die entsprechenden Ergebnisse sind entweder in Form von Odds Ratios (OR) oder Verhältnissen von Odds Ratios (ROR = ratio of odds ratios) dargestellt. Eine ROR = 1 bedeutet bei einem dichotomen Qualitätskriterium, dass die beiden OR gleich sind, Werte < 1 bedeuten dass die Studien, die das Kriterium erfüllt haben, weniger positive Ergebnisse (und dementsprechend die „schlechteren“ Studien optimistischere Ergebnisse) haben, und Werte > 1, dass Studien, die das Kriterium erfüllten, positivere Ergebnisse haben.

Ein möglicher Zusammenhang zwischen Studienqualität und –ergebnis wurde auf drei Weisen untersucht:

1. Prüfung einzelner Qualitätsaspekte (Komponentenanalyse): Für Studien, die ein definiertes Qualitätskriterium (Zuteilung explizit randomisiert, Zuteilung adäquat verblindet,

Doppelblindbedingungen, vollständiger Follow-up oder Intent-to-treat-Analyse) erfüllten oder nicht, wurden jeweils gepoolte Effektmaße (univariat und multivariat) berechnet.

2. Mindestscoreanalyse: Für Studien, die mindestens 3 Punkte beim Jadadscore, 5 Punkte auf der Internal Validity Scale oder beides („bessere Studien“) erreichten bzw. nicht erreichten, wurden gepoolte Effektmaße berechnet.
3. Kumulative Meta-Analyse: Die Primärstudien wurden konsekutiv entsprechend der erreichten Scorepunktzahl (für beide benutzte Skalen) in die Meta-Analyse einbezogen (d.h. zuerst nur die Studien mit der höchsten Punktzahl, dann nach und nach Studien mit niedrigeren Scorewerten).

3.2.3 Einfluss von Qualitätsmerkmalen auf Therapieerfolgswerten in Studien unterschiedlicher Designs am Beispiel Akupunktur bei chronischen Kopfschmerzen

3.2.3.1 Studienselektion

Eingeschlossen wurden Originalarbeiten, die folgendes Kriterium erfüllten: Behandlung von mindestens 5 Patienten mit idiopathischen oder chronischen Kopfschmerzen (Migräne, episodischer und chronischer Spannungskopfschmerz, Clusterkopfschmerzen, nicht genauer definierte chronische Kopfschmerzen) mit Akupunktur (Nadelinsertion an Akupunktur-, Schmerz- oder Triggerpunkten sowie andere Arten der Stimulation an definierten Akupunkturpunkten) zur Vorbeugung von Kopfschmerzattacken und Messung klinischer Zielparameter (z.B. Häufigkeit oder Intensität der Kopfschmerzen, Besserungsraten etc.). Nicht eingeschlossen wurden Studien zur Behandlung akuter Kopfschmerzattacken sowie Studien, die ausschließlich physiologische Parameter maßen. Studien in chinesischer Sprache konnten aufgrund mangelnder Ressourcen für eine Übersetzung nicht eingeschlossen werden.

3.2.3.2 Literatursuche

Um publizierte und unpublizierte Studien zu identifizieren, wurden Suchen in Medline (1966 bis Dezember 1998, Suchbegriffe: (acupuncture or acupressure or electroacupuncture) and (headache or migraine)), Embase (1989 bis Dezember 1998), dem Cochrane Controlled Trials Register (Ausgabe 1999/1), in individuellen Studiensammlungen und in Literaturverzeichnissen durchgeführt. Alle Referenzen und, soweit verfügbar, die entsprechenden Abstracts, wurden von 2 Reviewern durchgesehen. Kopien aller als potentiell relevant eingestuften Arbeiten wurden besorgt und formal auf die Erfüllung der Einschlusskriterien geprüft.

3.2.3.3 Extraktion und Beurteilung

Extraktion und Beurteilung erfolgten durch zwei Reviewer für die randomisierten Studien und durch einen für die nichtrandomisierten Studien. Mithilfe eines vorgetesteten Formulars wurden Informationen zur Quelle (bibliographische Angaben, Sprache, Herkunftsland, ob Medline-gelistet oder nicht), zu Patienten (Anzahl, Diagnosen, basale Einschlusskriterien, Kopfschmerzklassifikation, Art und Zahl der Prüfbereiche, Alter, Geschlecht, Dauer der Erkrankung, Ansprechen auf Vortherapien), zur Akupunktur (Strategie für Punktwahl, Zahl und Häufigkeit der Sitzungen) und zu Zielparametern extrahiert (Dauer des Follow-up, Definition von Therapieerfolg, Anteil Therapieerfolge nach Behandlung und im Follow-up bis und länger als 6 Monate, Zeitpunkt der frühesten Therapieerfolgsbeurteilung). Therapieerfolgsraten für die Meta-Analyse wurden wie folgt berechnet: Anzahl der Patienten mit Therapieerfolg zum frühesten verfügbaren Messzeitpunkt / Anzahl der mit Akupunktur behandelten Patienten einschließlich Abbrechern. Um zu untersuchen, ob nichtrandomisierte Studien relevante Zusatzinformationen erbringen, prüften wir, wieviele Studien a) einen Follow-up von mindestens 6 Monaten nach Therapie hatten, b) Daten zu mindestens 80% der ursprünglich behandelten Patienten berichten konnten, c) systematisch prognostische Faktoren untersuchten, d) Daten zu Nebenwirkungen und Komplikationen berichteten und e) in mehr als zwei Zentren durchgeführt worden waren und die Rekrutierungsstrategie (zur Beurteilung der Repräsentativität der Stichprobe) ausreichend detailliert beschrieben. Zusätzlich wurden anhand einer von White und Ernst (501) vorgeschlagenen Liste geprüft, ob die Akupunkturintervention ausreichend detailliert beschrieben worden war (Ergebnisse publiziert in (284)).

Die Studien wurden entsprechend ihres Designs dann klassifiziert in (explizit) randomisierte Studien und nichtrandomisierte Studien. Die nichtrandomisierten Studien wurden weiter unterteilt in a) nichtrandomisierte kontrollierte Studien, b) prospektive unkontrollierte Studien (d.h. Studien ohne Kontrollgruppe, die offensichtlich nach einem vorher festgelegten Protokoll durchgeführt wurden), c) Fallserien (unkontrollierte Studien ohne Evidenz für das Vorhandensein eines Studienprotokolls) und d) Befragungen bzw. Querschnittsstudien. Die Klassifikation erfolgte zweimal im Abstand von 5 Monaten durch den Autor.

Ziel der Qualitätsbeurteilung war, die *Angemessenheit von Beobachtung und Berichterstattung* – weitgehend unabhängig vom Studiendesign - zu überprüfen. Ob randomisiert oder nicht sollte eine in diesem Sinne „gute“ Studie dem Arzt, der Kopfschmerzpatienten therapiert, erlauben, sich ein ausreichend detailliertes Bild über die eingeschlossenen Patienten und deren Symptomatik über einen Zeitraum von mindestens 6 Monaten nach Therapie zu machen (unabhängig von der Frage, ob die Intervention die Veränderungen verursacht hat). Hierfür wurde eine ähnliche Kriterienliste wie bei den Übersichtsarbeiten zur Migräneprophylaxe mit Propranolol, Metoprolol und Flunarizin

verwendet (siehe 2.2.1.4.). 10 Items wurden abgefragt, die Antwortmöglichkeiten waren jeweils ja, nein und unklar:

1. Rekrutierung beschrieben? Hier mussten basale Angaben vorliegen, wo die Studie durchgeführt worden war (Zahl und Art der Zentren) und wie die Patienten rekrutiert wurden (Routinepatienten, überwiesene Patienten, durch Zeitungsanzeigen etc.).
2. Klare Kopfschmerzdiagnose? Es mußte angegeben sein, nach welcher Klassifikation bzw. nach welchen Kriterien die Kopfschmerzdiagnose gestellt worden war.
3. Patienten ausreichend charakterisiert? Angaben mindestens zu Alter, Geschlecht Erkrankungsdauer und ein Maß zum Schweregrad der Symptomatik.
4. Mindestens 4 Wochen Baselineperiode?
5. Mindestens 2 klinisch relevante Zielkriterien?
6. Verwendung eines Kopfschmerztagebuchs?
7. Kointerventionen beschrieben? Hier mußte angegeben sein, mit welchen Maßnahmen akute Kopfschmerzattacken behandelt wurden und wie sich dies im Verlauf veränderte.
8. Mindestens 90% der Patienten mit Daten bei Behandlungsende? Das Kriterium wurde als erfüllt gewertet, wenn bei Behandlungsende von mindestens 90% der Patienten Angaben zu den klinisch relevanten Zielkriterien vorlagen. Eine intent-to-treat-Analyse bei einer Abbruchrate von mehr als 10% wurde hier nicht anerkannt, da hier eine Beurteilung des tatsächlichen Verlaufs bei den Patienten nicht wirklich möglich ist, sondern nur geschätzt wird.
9. Mindestens 80% der Patienten mit Daten im Kurz-Follow-up (weniger als 6 Monate)?
10. Mindestens 80% der Patienten mit Daten im Langzeit-Follow-up (mehr als 6 Monate)?

3.2.3.4 Auswertung

Unterschiede zwischen randomisierten und nichtrandomisierten Studien bzgl. Studienmerkmalen und Erfüllung der Qualitätskriterien wurden mithilfe des Chi-quadrat- und des Mann Whitney Tests untersucht (keine Adjustierung für multiples Testen). Therapieerfolgswahrscheinlichkeiten für die Metaanalyse wurden mit einem allgemeinen linearen Regressionsmodell (random effects) berechnet (Funktion GLIMMIX in SAS). Es erfolgten sowohl univariate als auch multivariate Analysen.

3.3 Ergebnisse

3.3.1 Methodische Qualität randomisierter Studien in den Bereichen Akupunktur, Phytotherapie und Homöopathie

Die eingeschlossenen Studien zu Akupunktur, Phytotherapie und Homöopathie unterschieden sich in verschiedenen Aspekten (siehe Tabelle 5). Akupunkturstudien hatten im allgemeinen besonders geringe Fallzahlen (mediane Fallzahl in Asthmastudien 15 und in Kopfschmerzstudien 33), die Studien zu Echinacea und Hypericum waren fast ausnahmslos in Deutschland durchgeführt und publiziert worden. Abgesehen vom Bereich Akupunktur war weniger als die Hälfte der Studien in englischer Sprache verfügbar und über die Datenbank Medline identifizierbar. In allen drei Therapierichtungen war die große Mehrheit der Studien placebokontrolliert (bzw. Vergleich mit Scheinakupunktur).

Obwohl die Studienqualität stark variierte, ist festzuhalten, dass die Mehrheit der Studien deutliche Qualitätsmängel bzw. Mängel in der Berichterstattung offenbarte (siehe Tabelle 6). Nur ein geringer Teil der Studien beschrieb die Generierung der Randomisierungssequenz, die Verblindung der Zuteilung (allocation concealment) sowie die Zahl und die Gründe für Studienabbrüche und Ausschlüsse. Zwischen den drei Therapien gab es zum Teil erhebliche Unterschiede bzgl. der einzelnen Kriterien. So beschrieben Homöopathiestudien seltener die Gruppenzuteilung als randomisiert, während Probleme mit der Verblindung hier (verständlicherweise angesichts des oft typischen Geschmacks und Geruchs von Pflanzenextrakten sowie der generellen Probleme bzgl. eines Akupunkturplacebos) geringer waren. Phytotherapiestudien (dabei insbesondere die Hypericumstudien) beschrieben öfter Zahl und Gründe für Studienabbrüche und –ausschlüsse in ausreichender Weise. Insgesamt erreichten die Hypericumstudien deutlich höhere Scorewerte bei der Bewertung mit der Jadadskala (Mittelwert 3,56), sodass die Phytotherapiestudien hier insgesamt besser abschnitten als die Studien zu Akupunktur und Homöopathie.

Tabelle 5:
Merkmale der eingeschlossenen Studien

	Akupunktur	Phytotherapie	Homöopathie
Anzahl	36	42	129
Interventionen	Alle Akupunktur- interventionen	Echinacea Hypericum	Alle homöopath. Interventionen
Erkrankungen	Asthma Chron. Kopfschmerzen	Erkältung Depression	Keine Einschränkung
Fallzahlen (Median, Bereich)	28 (10-150)	100 (28-646)	60 (5-1306)
Art der Kontrollgruppe*			
- Placebo	29 (81%)	29 (69%)	114 (88%)
- Andere Therapie	9 (25%)	10 (24%)	19 (15%)
- Keine Therapie	2 (11%)	3 (7%)	5 (4%)
Herkunftsland			
- USA/Kanada	2 (6%)	-	4 (3%)
- Großbritannien	8 (22%)	1 (2)%	26 (20%)
- Deutschland	8 (22%)	39 (93%)	43 (33%)
- Frankreich	2 (6%)	-	33 (26%)
- sonstiges Europa	9 (25%)	2 (5%)	17 (13%)
- andere Länder	7 (19%)	-	6 (5%)
Verfügbar in englischer Sprache	25 (69%)	12 (29%)	49 (38%)
Gelistet in Medline	25 (69%)	7 (17%)	30 (23%)

*Einige Studien mit mehr als einer Kontrollgruppe

Tabelle 6:
Ergebnisse zur Bewertung der methodischen Qualität. Angegeben sind – wenn nicht anders beschrieben – die jeweilige Anzahl von Studien und in Klammern Prozentwerte

Kriterium (Scorepunkte)	Akupunktur	Phytotherapie	Homöopathie	P-Wert
Jadadscore				
Randomisation				
- nicht randomisiert/unklar (0)	4 (11%)	5 (12%)	45 (35%)	
- als randomisiert beschrieben (1)	29 (81%)	24 (57%)	63 (49%)	
- zusätzlich Beschreibung der Sequenzgenerierung (2)	3 (8%)	13 (31%)	21 (16%)	<0,001
Doppelblindbedingungen				
- nicht doppelblind/unklar (0)	17 (47%)	7 (17%)	24 (19%)	
- als doppelblind bezeichnet (1)	8 (22%)	18 (43%)	51 (39%)	
- zusätzlich glaubwürdig beschrieben (2)	11 (31%)	17 (40%)	54 (42%)	0,006
Studienabbrüche und Ausschlüsse				
- nicht oder unzureichend beschrieben (0)	22 (61%)	17 (41%)	96 (74%)	
- adäquat beschrieben (1)	14 (39%)	25 (59%)	33 (26%)	<0,001
Mittelwert (Standardabweichung) Jadadsummenscore	2,19 (1,17)	3,12 (1,33)	2,33 (1,36)	0,002
- Studien mit 0-2 Punkten	22 (61%)	12 (29%)	78 (60%)	
- Studien mit 3 und mehr Punkten	14 (39%)	30 (71%)	51 (40%)	0,001
Verblindung der Gruppenzuteilung				
- inadäquat (z.B. Alternierung)/unklar	31 (88%)	31 (74%)	88 (68%)	
- möglicherweise inadäquat (z.B. versiegelter	2 (6%)	-	-	
- adäquat (z.B. numerierte, neutrale Med.packungen)	2 (6%)	11 (26%)	41 (32%)	0,001
P-Werte aus χ^2 -Test				

Eine Publikation in englischer Sprache, in einer Medline-gelisteten Zeitschrift, nach 1990 und eine größere Fallzahl waren – bei der Auswertung über alle drei Therapien - mit signifikant besseren Jadadscores assoziiert (siehe Tabelle 7). Bei der Prüfung in Bezug auf die einzelnen Therapien waren die Unterschiede aufgrund der zum Teil kleinen Studienzahlen in einer bestimmten Kategorie nicht immer signifikant.

Tabelle 7:
Adjustierte und unadjustierte Jadadscores stratifiziert nach Publikationsart, -zeit, -sprache und Fallzahl (Adjustierung für die jeweils anderen Faktoren mithilfe von linearer Regression).

	Akupunktur			Phytotherapie			Homöopathie			Alle ^a		
	Mittlerer Jadadscore			Mittlerer Jadadscore			Mittlerer Jadadscore			Mittlerer Jadadscore		
	n	Adjust	Unadj	n	Adjust	Unadj	n	Adjust	Unadj	n	Adjust	Unadj
Publikation in												
- Medline-gelisteter Zeitschrift	25	2,42	2,40	9	3,98	4,56	30	3,10	3,27	46	2,95	3,04
- andere Quellen	11	1,68	1,73	32	2,88	2,72	99	2,10	2,05	74	2,42	2,36
P-Wert		0,115			0,030			0,000			0,031	
Publikation												
- vor 1990	24	2,00	2,08	14	2,66	2,21	85 ^b	2,14	2,07	68	2,22	2,21
- 1990 und später	12	2,58	2,42	27	3,63	3,59	42	2,74	2,88	52	3,15	3,17
P-Wert		0,176			0,084			0,009			0,000	
Verfügbar												
- in Englisch	25	2,33	2,40	12	3,61	4,25	49	2,70	2,84	61	2,88	2,97
- nur in anderer Sprache	11	1,89	1,73	29	2,92	2,66	80	2,11	2,03	59	2,36	2,27
P-Wert		0,353			0,148			0,009			0,027	
Fallzahl												
< 100	35	Keine	Keine	19	3,13	3,16	91	2,19	2,14	87	2,47	2,48
≥ 100	1	Ana-	Ana-	22	3,11	3,09	38	2,67	2,79	33	3,03	3,00
P-Wert		lyse	lyse		0,947			0,041			0,023	

^aUm zu verhindern, dass die überproportional große Zahl Homöopathiestudien das Ergebnis beeinflusst, wurden nur 42 zufällig ausgewählte Homöopathiestudien in die Analysen einbezogen

^bZwei unpublizierte Studien ohne Datum

n = Anzahl Studien; Adjust = adjustierte Scores; Unadj = unadjustierte Scores

Doppelblindbedingungen, adäquate Verblindung der Gruppenzuteilung (allocation concealment) und eine vollständige Beschreibung von Studienabbrüchen und –ausschlüssen waren in neueren Studien signifikant häufiger. Bei den anderen Merkmalen war die Assoziation zwischen einzelnen Merkmalen und einzelnen Qualitätskriterien weniger konsistent (siehe Tabelle 8).

Tabelle 8:

Einfluss von Publikationssprache, -quelle und -jahr sowie der Fallzahl auf die Beschreibung von zentralen methodischen Qualitätskriterien. Angegeben sind die Odds Ratios (OR) mit 95%-Konfidenzintervallen (95%KI) für die Studien, bei denen ein Kriterium erfüllt war im Vergleich zu denen, die ein Kriterium nicht erfüllten (eine OR von 1,69 für in englischer Sprache publizierte Studien bzgl. Doppelblindbedingungen bedeutet z.B., dass diese Studien häufiger doppelblind waren als in anderen Sprachen publizierte Studien)

Merkmal	Doppelblindbedingungen		Angemessenes Concealment		Beschreibung Abbrecher	
	OR (95%KI)	p	OR (95%KI)	p	OR (95%KI)	p
Publiziert in Englisch	1,69 (0,64-4,47)	0,290	4,54 (1,47-14,02)	0,009	0,92 (0,40-2,31)	0,921
Medline-gelistete Publikation	0,77 (0,28-2,06)	0,600	0,70 (0,24-2,03)	0,511	3,50 (1,41-8,68)	0,007
Seit 1990 publiziert	3,72 (1,37-10,10)	0,010	2,57 (1,02-6,46)	0,046	3,39 (1,51-7,59)	0,003
Fallzahl ≥ 100	2,09 (0,67-6,26)	0,208	4,64 (1,70-12,65)	0,003	1,32 (0,54-3,31)	0,523

3.3.2 Einfluss von Aspekten der methodischen Qualität auf das Studienergebnis am Beispiel der placebokontrollierten Studien zur Homöopathie

Im Durchschnitt erhielten die 89 einbezogenen Studien 2,58 (Standardabweichung 1,29) von maximal 5 möglichen Punkten im Jadadscore und 4,20 (1,46) von 7 möglichen bei der Internal Validity Scale. Vierzig Studien (45%) erreichten die als Mindestscore für „bessere“ Studien vorher festgelegten 3 oder mehr Punkte im Jadadscore, 34 Studien (38%) den Mindestscore (5 von 7 Punkten) auf der Internal Validity Scale und 26 (29%) in beiden Skalen. 64 Studien (72%) waren explizit randomisiert, bei 21 (24%) war die genaue Zuteilungsmethode nicht beschrieben und 4 Studien (4%) waren quasi-randomisiert (Alternierung, Zuteilung nach Geburts- oder Aufnahmedatum). 34 Studien (38%) beschrieben eine angemessene Methode zur Verblindung der Gruppenzuteilung (in der Regel Verwendung konsekutiv numerierter, identisch verpackter Prüfpräparate), 50 (56%) machten hierzu keine Angaben und in 5 Studien (6%) war die Gruppenzuteilung nicht angemessen verblindet. 81 Untersuchungen (91%) wurden unter Doppelblindbedingungen durchgeführt, 3 (5%) einfachblind und bei 5 (6%) wurden keine Angaben gemacht. 28 Studien (37%) bezogen alle Patienten in die Analyse ein, in 20 (22%) erschienen relevante Verzerrungen unwahrscheinlich, in 19 (21%) dagegen möglich oder wahrscheinlich und 22 (25%) machten ungenügende Angaben.

Die gepoolte Odds Ratio (OR) über alle 89 Studien war 2,45 (95%-Konfidenzintervall 2,05-2,93). Da die Effektgrößen der Studien ausserordentlich heterogen waren ($\tau^2 = 0,43$, 95%-KI 0,25-0,90; p

= $2,4 \times 10^{-13}$) werden im folgenden ausschließlich „random effects“ Analysen berichtet. Wenn nur die Studien gepoolt wurden, die explizit randomisiert, mit angemessen verblindeter Zuteilung oder doppelblind waren, ergaben sich ORs von 2,23 (9% Verminderung des Effekts im Vergleich zu dem gepoolten Maß über alle 89 Studien), 2,00 (18% Verminderung) bzw. 2,18 (11% Verminderung; siehe auch Tabelle 9). Dass der Ausschluss von nur 8 nicht doppelblinden Studien zu einer Verminderung der OR um 11% führte, zeigt, dass diese Studien wesentlich größere Effekte berichteten als doppelblinde. Der Ausschluß verminderte ausserdem die Heterogenität deutlich (siehe Tabelle 9, letzte Spalte). Studien mit vollständigem Follow-up bzw. intent-to-treat Analyse berichteten etwas größere Effekte als Studien, bei denen dieses Kriterium nicht erfüllt war. Das Kriterium Doppelblindbedingungen hatte sowohl in der uni- als auch in der multivariaten Analyse den größten Einfluss. Das Verhältnis der ORs (ROR) nicht-doppelblinder vs. doppelblinder Studien war 0,24 (95%-KI 0,12-0,46; $p < 0,0001$) bzw. 0,26 (95%-KI 0,14-0,51; $p = 0,0002$). In der multivariaten Analyse erwies sich auch der Einfluß einer explizit randomisierten Zuteilung als signifikant (ROR = 0,64; 95%-KI 0,43-0,94; $p = 0,03$). Verblindung der Zuteilung und vollständiger Follow-up bzw. intent-to-treat Analyse hatten keinen signifikanten Einfluss. Die multivariate Analyse einzelner Qualitätskomponenten führte im Vergleich zu den anderen Analysen auch zur stärksten Reduzierung der Heterogenität.

Tabelle 9:
Auswirkungen von Qualitätsindikatoren auf das Studienergebnis: Analysen einzelner Kriterien (uni- und multivariat) und Mindestscoreanalyse

	Kriterium erfüllt		Kriterium nicht erfüllt			
	n	OR (95%KI)	n	OR (95%KI)	ROR (95%KI)	Heterogenität
Analyse einzelne Kriterien (univariat)						
Explizit randomisiert	64	2,23 (1,81-2,75)	25	3,40 (2,32-4,97)	0,66 (0,43-1,01)	0,41 (0,23-0,87)
Angemessene Zuteilungsverblindung	34	2,00 (1,50-2,65)	55	2,83 (2,23-3,58)	0,71 (0,49-1,02)	0,39 (0,22-0,87)
Doppelblindbedingungen	81	2,18 (1,83-2,60)	8	9,14 (4,81-17,4)	0,24 (0,12-0,46)	0,31 (0,18-0,71)
Vollständige Analyse	28	3,03 (2,12-4,33)	61	2,29 (1,84-2,86)	1,31 (0,88-2,00)	0,44 (0,26-0,91)
Analyse einzelne Kriterien (multivariat)						
Explizit randomisiert		2,15 (1,78-2,59)		3,36 (2,37-4,78)	0,64 (0,43-0,94)	0,28 (0,15-0,69)
Angemessene Zuteilungsverblindung		2,11 (1,62-2,74)		2,50 (2,00-3,11)	0,84 (0,60-1,18)	
Doppelblindbedingungen		2,08 (1,75-2,47)		7,85 (4,10-15,0)	0,26 (0,14-0,51)	
Vollständige Analyse		2,70 (1,97-3,72)		2,21 (1,81-2,69)	1,23 (0,85-1,77)	
Mindestscoreanalyse						
Jadadscore > 2	40	1,81 (1,41-2,32)	49	3,23 (2,53-4,13)	0,56 (0,40-0,79)	0,34 (0,19-0,78)
Internal Validity Score > 4,5	34	1,97 (1,50-2,59)	55	2,92 (2,29-3,73)	0,67 (0,47-0,97)	0,40 (0,23-0,86)
Beide Kriterien erfüllt	26	1,72 (1,28-2,31)	63	2,95 (2,37-3,67)	0,58 (0,40-0,79)	0,36 (0,20-0,82)

n = Anzahl Studien, OR = Odds ratio, ROR = Ratio der Odds Ratios, 95%KI = 95%-Konfidenzintervall

Wenn ausschließlich Studien mit 3 oder mehr Punkten im Jadadscore, 5 oder mehr Punkten auf der Internal Validity Scale bzw. mit Erfüllung beider Voraussetzungen in die Meta-Analyse eingingen (Mindestscoreanalyse), ergaben sich ORs von 1,81 (95%-KI 1,41-2,32, 28% Verminderung), 1,95 (95%-KI 1,50-2,59, 22% Verminderung) bzw. 1,72 (1,28-2,31, 30% Verminderung).

Wenn die Studien konsekutiv entsprechend der erreichten Punktzahl auf der Internal Validity Scale in die Meta-Analyse eingingen (kumulative Meta-Analyse), zeigte sich ein Trend zu größeren Effekten, wenn qualitativ „schlechtere“ Studien einbezogen wurden (siehe Tabelle 10, oberer Teil). Bei der entsprechenden Analyse auf Basis des Jadadscores zeigten die besten Studien (5 Scorepunkte) dagegen größere Effekte als Studien mit 3 oder 4 Scorepunkten (Tabelle 10, unterer Teil).

Tabelle 10:
Auswirkungen von Qualitätsindikatoren auf das Studienergebnis: Kumulative Meta-Analyse

Scorepunktzahl (Studienzahl)	OR (95%KI) der Studien mit der entsprechenden Scorepunktzahl	Scorepunktzahl (kumulative Studienzahl)	Kumulative OR (95%KI)
Internal Validity Scale			
7,0 (5)	1,55 (0,77-3,10)	= 7,0 (5)	1,55 (0,77-3,10)
6,5 (2)	4,61 (0,91-23,4)	≥ 6,5 (7)	2,02 (1,06-3,85)
6,0 (4)	1,35 (0,74-2,45)	≥ 6,0 (11)	1,67 (1,10-2,54)
5,5 (13)	1,74 (1,38-2,19)	≥ 5,5 (24)	1,71 (1,36-2,15)
5,0 (10)	2,95 (2,00-4,34)	≥ 5,0 (34)	1,89 (1,53-2,35)
4,5 (9)	4,33 (1,86-10,1)	≥ 4,5 (43)	2,09 (1,69-2,59)
4,0 (13)	3,16 (2,00-4,98)	≥ 4,0 (56)	2,29 (1,88-2,80)
3,5 (9)	2,82 (1,57-5,05)	≥ 3,5 (65)	2,35 (1,94-2,83)
3,0 (8)	1,97 (1,14-3,40)	≥ 3,0 (73)	2,29 (1,92-2,74)
2,5 (5)	3,91 (0,97-15,8)	≥ 2,5 (78)	2,42 (2,00-2,91)
2,0 (9)	2,47 (1,11-5,52)	≥ 2,0 (87)	2,44 (2,02-2,93)
1,0 (2)	6,92 (1,53-31,4)	alle (89)	2,47 (2,06-2,97)
Jadadscore			
5 (10)	2,00 (1,37-2,91)	= 5 (10)	2,00 (1,37-2,91)
4 (11)	1,42 (1,02-1,99)	≥ 4 (21)	1,68 (1,29-2,18)
3 (19)	1,83 (1,44-2,34)	≥ 3 (40)	1,73 (1,44-2,08)
2 (32)	3,24 (2,27-4,63)	≥ 2 (72)	2,26 (1,87-2,73)
1 (15)	3,58 (2,11-6,08)	≥ 1 (87)	2,44 (2,02-2,93)
0 (2)	6,92 (1,53-31,4)	alle (89)	2,47 (2,06-2,97)

OR = Odds ratio, 95%KI = 95%-Konfidenzintervall

Wenn die Odds Ratios und Qualitätsscores der einzelnen Studien in einem Scatterplotdiagramm aufgetragen werden (ohne Abbildung), wird jedoch deutlich, dass kein klarer linearer Zusammenhang zwischen diesen beiden Variablen besteht. D.h. zwischen den Studien mit einer bestimmten Scorepunktzahl streuen die berechneten Effektgrößen in hohem Maße.

3.3.3 Einfluss von Qualitätsmerkmalen auf Therapieerfolgswerten in Studien unterschiedlicher Designs am Beispiel Akupunktur bei chronischen Kopfschmerzen

Von 97 identifizierten Studien zur Akupunktur bei Kopfschmerzpatienten erfüllten 59 die Einschlusskriterien. Der häufigste Ausschlussgrund war die Vermischung der Daten von Kopfschmerz- und anderen Schmerzpatienten. 24 der eingeschlossenen Studien waren randomisiert und 35 nichtrandomisiert (5 nichtrandomisierte kontrollierte Studien, 10 prospektive unkontrollierte Studien, 10 Fallserien und 10 Befragungen). Die Studien waren sehr heterogen bzgl. Patienten, Akupunktur, Art und Zeitpunkten für die Ergebnismessung. Die Tabellen 11 und 12 fassen wesentliche Informationen zu allen 59 Studien zusammen.

Tabelle 11:
Allgemeiner Überblick der eingeschlossenen Studien zur Akupunktur bei chronischen Kopfschmerzen: 1. Randomisierte Studien

Erstautor	Art	Qual.-	Definition	Behandelte	Therapieerfolg (%)		
	Kopf- Schmerz	kriterien erfüllt			Therapieerfolg	Patienten	Nach Follow-up
					Behandl.	< 6 Mon.	≥ 6 Mon.
Randomisierte Studien							
Ahonen (5)	Spannung	4	> 2 Monate Besserung*	12	k.A.	58	k.A.
Baust (32)	Migräne	4	>50% Index-Reduktion	23	61	k.A.	k.A.
Carlsson (57)	Spannung	5	-	30	k.A.	k.A.	k.A.
Ceccherelli (59)	Migräne	5	>50% Index-Reduktion	15	87	k.A.	k.A.
Doerr-Proske (94)	Migräne	7	-	10	k.A.	k.A.	k.A.
Dowson (95)	Migräne	8	> 50% Häufigkeits-Reduktion	25	32	k.A.	k.A.
Hansen (162)	Spannung	2	> 33% Index-Reduktion	25	28	k.A.	k.A.
Henry (179)	Migräne	5	Globalurteil (Patient)	20	65	k.A.	k.A.
Hesse (181)	Migräne	8	-	42	k.A.	k.A.	k.A.
Heydenreich (184)	Migräne	3	Unklar	20	k.A.	k.A.	80
Heydenreich (183)	Migräne	3	Unklar	50	k.A.	k.A.	80
Johansson (215)	Spannung	0	-	17	k.A.	k.A.	k.A.
Kubiena (252)	Migräne	4	> 33% Index-Reduktion	15	75	k.A.	k.A.
Lehmann (272)	Migräne	3	-	21	k.A.	k.A.	k.A.
Lenhard (275)	Migräne	7	> 33% Häufigkeits-Reduktion	16	k.A.	56	k.A.
Loh (292)	verschied.	3	Globalurteil (Patient)	23	35	k.A.	k.A.
Okazaki (348)	Migräne	1	Unklar	20	75	k.A.	k.A.
Pintov (368)	Migräne	7	-	11	k.A.	k.A.	k.A.
Tavola (447)	Spannung	10	> 50% Index-Reduktion	15	53	k.A.	40
Vincent (478)	Migräne	9	> 33% Häufigkeits-Reduktion	16	47	k.A.	k.A.
Vincent (479)	Spannung	9	> 50% Index-Reduktion	14	64	43	k.A.
Weinschütz (494)	Migräne	5	stat. Einzelfallanalyse	20	50	65	k.A.
Weinschütz (495)	Migräne	5	stat. Einzelfallanalyse	20	45	75	k.A.
White (500)	Spannung	5	-	5	k.A.	k.A.	k.A.

k.A. = keine Angabe; *Operationalisierung unklar

Tabelle 12:
Allgemeiner Überblick der eingeschlossenen Studien zur Akupunktur bei chronischen Kopfschmerzen: Nichtrandomisierte Studien

Erstautor	Art	Qual.-	Definition	Behandelte	Therapieerfolg (%)		
	Kopf- Schmerz	kriterien Erfüllt			Nach Follow-up		
			Therapieerfolg	Patienten	Behandl.	< 6 Mon.	≥ 6 Mon.
Nichtrandomisierte kontrollierte Studien							
Jensen (209)	unklar	8	unklar	32	79	k.A.	k.A.
Guccione (158)	Migräne	1	-	24	k.A.	k.A.	k.A.
Sold-Darseff (416)	Migräne	1	unklar	30	93	k.A.	k.A.
Turk (463)	Migräne	1	unklar	100	86	k.A.	k.A.
Zibin (529)	Migräne	0	unklar	340	99	k.A.	k.A.
Unkontrollierte prospektive Studien							
Baischer (24)	Migräne	10	> 33% Index-Reduktion	31	70	k.A.	58
Boivie (42)	Migräne	10	Unklar	30	k.A.	56	56
Jensen (208)	Spannung	6	Unklar	21	67	57	k.A.
Jensen (210)	Spannung	5	Unklar	16	63	k.A.	k.A.
Ceccherelli (61)	verschied.	4	> 50% Index-Reduktion	12	100	k.A.	k.A.
Ceccherelli (60)	verschied.	6	> 40% Index-Reduktion	92	76	k.A.	k.A.
Kim (236)	Migräne	3	Besserung*	25	k.A.	80	k.A.
Major (305)	Migräne	8	> 50% Häufigkeits-Reduktion	33	57	k.A.	k.A.
Tolksdorf (462)	verschied.	5	> 50% Index-Reduktion	70	59	k.A.	38
Zwölfer (532)	Spannung	3	Globalurteil (Patient)	20	80	k.A.	k.A.
Fallserien							
Batra (31)	Migräne	3	> 50% Besserung*	20	k.A.	65	k.A.
Cheng (68)	verschied.	2	Keinerlei Kopfschmerz	33	91	k.A.	k.A.
Fischer (133)	unklar	2	> 50% Symptom-Reduktion*	189	75	k.A.	k.A.
Heydenreich (183)	Migräne	0	> 50% Besserung *	320	k.A.	k.A.	88
Heydenreich (182)	Migräne	3	> 50% Besserung *	296	k.A.	k.A.	76
Laitinen (260)	Migräne	5	Besserung*	39	92	k.A.	46
Lawrence (267)	verschied.	1	Besserung*	96	92	k.A.	k.A.
Spoerel (429)	verschied.	2	Häufigkeits- und Intens.red.*	38	k.A.	84	k.A.
Stux (444)	Migräne	0	Unklar	50	88	k.A.	k.A.
Yamauchi (525)	verschied.	2	Unklar	7	71	k.A.	k.A.
Befragungen							
Carlsson (56)	verschied.	4	Globalurteil (Patient)	79	57	k.A.	23
Chen (67)	unklar	1	Globalurteil (Patient)	44	k.A.	k.A.	68
Haug (171)	unklar	1	Globalurteil (Patient)	49	k.A.	k.A.	84
Junilla (223)	unklar	2	> 30% Symptom-Reduktion*	115	k.A.	k.A.	77
Junilla (224)	verschied.	2	Globalurteil (Patient)	69	81	k.A.	k.A.
Kubiena (251)	Migräne	1	Globalurteil (Patient)	118	k.A.	62	k.A.
Maxion (311)	Migräne	1	Globalurteil (Patient)	13	15	k.A.	k.A.
McKibbin (315)	verschied.	4	-	107	k.A.	k.A.	k.A.
Pöntinen (377)	Migräne	2	Globalurteil (Patient)	107	k.A.	k.A.	57
Shifman (408)	verschied.	0	Globalurteil (Patient)	33	58	k.A.	k.A.

k.A. = keine Angabe; *Operationalisierung unklar

Randomisierte Studien hatten signifikant kleinere Fallzahlen als nichtrandomisierte (siehe Tabelle 13). Insgesamt wurden in den randomisierten Studien 535 Patienten mit Akupunktur behandelt, in den nichtrandomisierten 2695. In den randomisierten Studien war die Diagnose häufiger explizit Migräne; außerdem wurden die Untersuchungen etwas häufiger in spezialisierten Zentren durchgeführt. Alter, Geschlecht und Erkrankungsdauer waren dagegen sehr ähnlich. Nur 14 Studien (7 randomisierte und 7 nichtrandomisierte) machten Aussagen darüber, wie die Patienten auf frühere Versuche zur Kopfschmerzprophylaxe angesprochen hatten: in jeweils 4 Studien hatten sich die Vortherapien als unwirksam erwiesen. Die Akupunktursitzungen dauerten in den nichtrandomisierten Studien signifikant länger, während die Sitzungszahl ähnlich war. In den randomisierten Studien war die Wahl der Akupunktur ähnlich oft individualisiert oder weitgehend individualisiert wie in den nichtrandomisierten Studien.

Tabelle 13:
Vergleich von randomisierten und nichtrandomisierten Studien zur Akupunktur bei chronischen Kopfschmerzen: Merkmale von Patienten und Interventionen

	Randomisiert (n = 24)	Nichtrandomisiert (n = 35)	p-Wert
Patientenmerkmale			
Patienten behandelt mit Akupunktur (Median, Bereich)	20 (5-100)	39 (7-340)	<0,001
Diagnose			
- Migräne	16 (67%)	16 (46%)	
- Spannungskopfschmerz	7 (29%)	3 (8%)	
- verschiedene/unklar	1 (4%)	16 (46%)	0,002
Durchschnittsalter (Median, Bereich)	40 (32-46; mi: 7)	40 (22-53, mi: 19)	0,515
%-Anteil Frauen (Median, Bereich)	80 (57-100, mi: 3)	78 (33-100, mi: 10)	0,667
Erkrankungsdauer > 5 Jahre	18 (94%, mi: 6)	18 (100%, mi: 17)	0,866
Setting	mi: 5	mi: 6	
- Schmerzzambulanzen u.ä.	17 (89%)	20 (69%)	
- Praxen	2 (11%)	9 (31%)	0,161
Interventionsmerkmale			
Anzahl Akupunktursitzungen (Median; Bereich)	8 (4-14; mi: 3)	10 (1-30; mi: 11)	0,731
Sitzungsdauer in Minuten (Median; Bereich)	15 (1-30; mi: 7)	22.5 (10-45; mi: 10)	0,001
Individualisierte Intervention	9 (43%, mi: 3)	14 (48%, mi: 6)	0,631

mi = missing (Anzahl der Studien ohne Angaben)

In den randomisierten Studien waren die Kriterien für die Angemessenheit von Beobachtung und Berichterstattung im Durchschnitt deutlich häufiger erfüllt wie in den nichtrandomisierten Studien (siehe Tabelle 14). Die Anzahl der erfüllten Kriterien lag jedoch für beide Studiengruppen zwischen 0 und 10, was deutlich macht, dass jeweils indiskutabel berichtete sowie hochwertige

Untersuchungen vorlagen. 10 der 24 randomisierten und 26 der 35 nichtrandomisierten Studien erfüllten weniger als 5 Kriterien. Randomisierte Studien machten signifikant häufiger klare Angaben zur Kopfschmerzdiagnose, hatten häufiger eine vierwöchige Baselinephase, mindestens zwei klinisch relevante Zielkriterien und benutzten signifikant häufiger Kopfschmerztagebücher zur Beurteilung. Rekrutierung und Kointerventionen (Behandlung der akuten Kopfschmerzen) wurden unabhängig vom Studientyp fast nie befriedigend beschrieben. Die Follow-up-Perioden waren häufig kurz oder die Zahl der Patienten mit fehlenden Informationen hoch.

Insgesamt 50 Studien präsentierten Angaben zum Therapieerfolg zumindest zu einem der drei vordefinierten Zeitperioden (nach Behandlung, Follow-up bis und über 6 Monate). Randomisierte Studien machten häufiger Angaben darüber, wie der Therapieerfolg definiert war. Die Definitionen waren in vielen Fällen jedoch unterschiedlich. Die gepoolte Therapieerfolgsrate war 59% (95%-Konfidenzintervall 48-69%) in den randomisierten und 78% (72-83%, $p < 0,001$) in den nichtrandomisierten Studien.

Tabelle 14:

Vergleich von randomisierten und nichtrandomisierten Studien zur Akupunktur bei chronischen Kopfschmerzen: Angemessenheit von Beobachtung und Berichterstattung, Zielkriterien und Therapieerfolgsraten

Studiendesign (n)	Randomisiert (24)	Nichtrandomisiert (35)	p-Wert
Angemessenheit von Beobachtung und Berichterstattung			
Rekrutierung beschrieben	8 (33%)	9 (26%)	0,559
Klare Kopfschmerzdiagnose	16 (67%)	8 (23%)	0,001
Patienten ausreichend charakterisiert	17 (71%)	17 (49%)	0,112
Mindestens 4 Wochen Baseline	15 (63%)	5 (14%)	<0,001
Kointerventionen beschrieben	7 (29%)	10 (29%)	1,000
Mindestens 2 klinisch relevante Zielkriterien	20 (83%)	18 (51%)	0,014
Verwendung von Kopfschmerztagebüchern	18 (75%)	8 (23%)	<0,001
≥ 90% der Patienten mit Daten nach Behandlung	11 (46%)	19 (54%)	0,601
≥ 80% der Patienten mit Daten im Follow-up < 6 Mon.	6 (25%)	10 (29%)	1,000
≥ 80% der Patienten mit Daten im Follow-up ≥ 6 Mon.	2 (8%)	6 (17%)	0,453
Mediane Zahl (Bereich) erfüllter Kriterien	5 (0-10)	2 (0-10)	0,008
Mindestens 5 Kriterien erfüllt	14 (58%)	9 (26%)	0,016
Subjektive Qualitätsbeurteilung "schlecht"	6 (25%)	29 (83%)	<0,001
Zielkriterien			
Angaben zum Therapieerfolg verfügbar	16 (67%)	33 (94%)	0,024
Angaben zum Therapieerfolg bis 3 Mon. nach Beh.	16 (89%)	21 (64%)	0,634
≥ 80% der Patienten mit Daten bei Erfolgsbestimmung	10 (56%)	21 (64%)	0,765
Therapieerfolgsrate (95%-Konfidenzintervall)	0,59 (0,48-0,69)	0,78 (0,72-0,83)	<0,001

10 (42%) randomisierte und 14 (40%) nichtrandomisierte Studien präsentierten Daten zu Follow-up-Perioden von mindestens 6 Monaten (siehe Tabelle 15). Jedoch in nur 2 (8%) der randomisierten und 6 (17%) der nichtrandomisierten Studien lagen zu diesen Zeitpunkten auch von mehr als 80% der behandelten Patienten Daten vor. In all diesen Studien war bei einem relevanten Anteil der Patienten die nach der Akupunktur beobachtete Verbesserung stabil geblieben. Nur 2 randomisierte (8%) bzw. 3 nichtrandomisierte (9%) Studien enthielten Analysen des Einflusses prognostischer Faktoren auf den Therapieerfolg, jedoch können die Ergebnisse aufgrund der kleinen Fallzahlen und der explorativen Analyseansätze bestenfalls hypothesengenerierenden Charakter haben. Alle 59 Studien waren unizentrisch; die Generalisierbarkeit ist aufgrund der schlechten Beschreibung der Rekrutierungsstrategien unklar, erscheint jedoch gering. Ergebnisse zu Nebenwirkungen und Komplikationen schließlich wurden nur in 4 randomisierten (17%) bzw. einer nichtrandomisierten (3%) Studie präsentiert.

Tabelle 15:
Vergleich von randomisierten und nichtrandomisierten Studien zur Akupunktur bei chronischen Kopfschmerzen: Information zu Langzeitergebnissen, prognostischen Faktoren, Generalisierbarkeit und Nebenwirkungen bzw. Komplikationen

Studiendesign (n)	Randomisierte (24)	Nichtrandomisierte (35)	p-Wert
Follow-up nach Behandlung mindestens 6 Monate	10 (42%)	14 (40%)	1,000
Ergebnisdaten zu mind. 80% der Patienten 6 Monate nach Behandlung	2 (8%)	6 (17%)	0,453
Analyse prognostischer Faktoren	2 (8%)	3 (9%)	1,000
Mind. 3 Studienzentren + transparente Rekrutierung	-	-	
Daten zu Nebenwirkungen und Komplikationen	4 (17%)	1 (3%)	0,148

In den univariaten Analysen hatten Diagnose, Geschlecht, Alter, Zahl und Dauer der Behandlungssitzungen, Art der Therapieerfolgsdefinition, Zeitpunkt der Erfolgsbestimmung und die Vollständigkeit der Daten bei der Analyse keinen signifikanten Einfluß auf die beobachteten Therapieerfolgsraten (Tabelle 16). Dagegen hatten Studien, die ihre Rekrutierungsstrategie beschrieben, klare Kopfschmerzdiagnosen referierten, mindestens 4 Wochen Baseline schrieben und Kopfschmerztagebücher benutzten, signifikant niedrigere Erfolgsraten. Abbildung 12 zeigt, dass sowohl ein Randomisationsdesign als auch eine höhere Anzahl erfüllter Kriterien bzgl. Angemessenheit von Beobachtung und Berichterstattung mit niedrigeren Erfolgsraten assoziiert waren. Wir waren nicht in der Lage, ein, adäquates Modell für die multiple Regressionsanalyse zu definieren. Außer dem Faktor Design (randomisiert oder nicht) hatten einzelne andere Faktoren kaum konsistente Effekte.

Tabelle 16:
Gepoolte Therapieerfolgsraten (random effects model) in allen Studien (randomisiert und nichtrandomisiert), die ein bestimmtes Kriterium nicht erfüllen bzw. erfüllen (univariate logistische Regression)

Kriterium	Therapieerfolgsrate (95%-Konfidenzintervall) Kriterium nicht erfüllt vs. Kriterium erfüllt	p-Wert
Patientenmerkmale		
Nur Migränepatienten	0,72 (0,63-0,79) vs. 0,73 (0,65-0,81)	0,7543
Interventionsmerkmale		
10 Behandlungssitzungen à mindestens 15 Minuten	0,76 (0,63-0,86) vs. 0,69 (0,60-0,78)	0,3414
Angemessenheit von Beobachtung und Berichterstattung		
Rekrutierung beschrieben	0,76 (0,71-0,82) vs. 0,62 (0,50-0,72)	0,0138
Klare Kopfschmerzdiagnose	0,78 (0,72-0,83) vs. 0,62 (0,52-0,72)	0,0037
Patienten charakterisiert	0,78 (0,70-0,84) vs. 0,68 (0,60-0,75)	0,0581
Mindestens 4 Wochen Baseline	0,77 (0,72-0,82) vs. 0,61 (0,49-0,72)	0,0099
Kointerventionen beschrieben	0,74 (0,67-0,80) vs. 0,69 (0,57-0,79)	0,4695
Mindestens 2 klinisch relevante Zielkriterien	0,77 (0,68-0,83) vs. 0,70 (0,62-0,77)	0,2256
Verwendung eines Kopfschmerztagebuchs	0,78 (0,73-0,83) vs. 0,61 (0,51-0,70)	0,0015
≥ 90% der Patienten mit Daten nach Behandlung	0,75 (0,68-0,82) vs. 0,69 (0,60-0,77)	0,2208
≥ 80% der Patienten mit Daten im Follow-up < 6 Mon.	0,75 (0,68-0,80) vs. 0,66 (0,54-0,77)	0,1735
≥ 80% der Patienten mit Daten im Follow-up ≥ 6 Mon.	0,73 (0,66-0,79) vs. 0,72 (0,58-0,82)	0,8816
Mindestens 5 Kriterien erfüllt	0,77 (0,71-0,82) vs. 0,63 (0,53-0,73)	0,0186
Merkmale der Zielkriterien		
Erfolgsdefinition ≥ 50% Reduktion	0,74 (0,68-0,80) vs. 0,70 (0,57-0,80)	0,4966
Erfolgsbestimmung < 3 Monate nach Therapieende	0,71 (0,64-0,77) vs. 0,79 (0,66-0,88)	0,2383
≥ 80% der Patienten mit Daten bei Erfolgsbestimmung	0,75 (0,66-0,82) vs. 0,71 (0,63-0,78)	0,4425

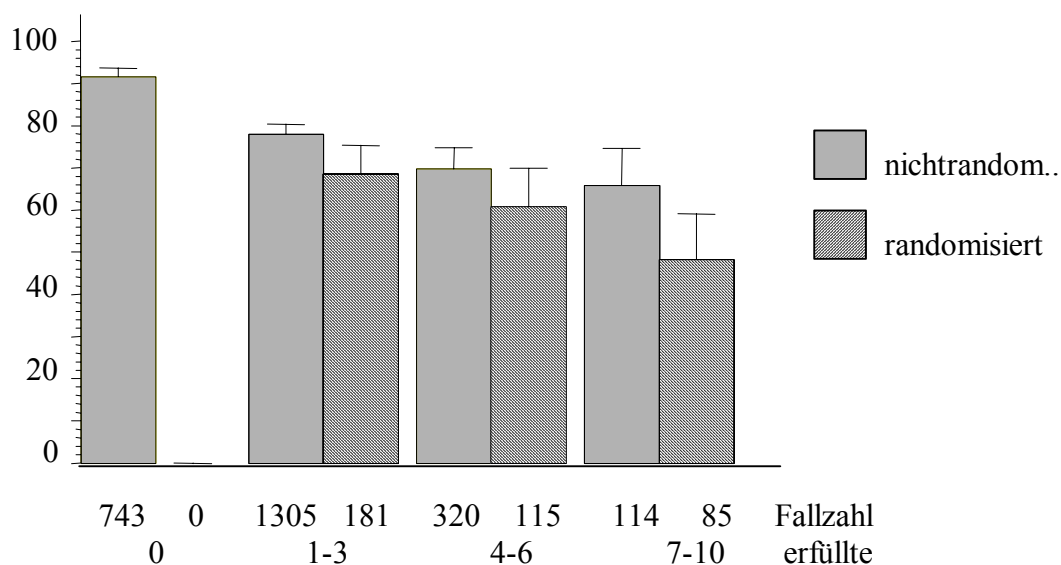


Abbildung 12:
Therapieerfolgsraten (in % mit 95%-Konfidenzintervallen) in randomisierten und nichtrandomisierten Studien zur Akupunktur bei chronischen Kopfschmerzen in Abhängigkeit von der Zahl erfüllter Kriterien in Bezug auf Angemessenheit der Beobachtung und Berichterstattung

3.4 Diskussion

3.4.1 Methodische Qualität randomisierter Studien in den Bereichen Akupunktur, Phytotherapie und Homöopathie

Die Ergebnisse in Abschnitt 3.3.1. zeigen, dass (1) viele randomisierte Studien in den Bereichen Phytotherapie, Homöopathie und Akupunktur relevante methodische Schwächen haben, dass (2) Studiencharakteristika wie Fallzahl und die spezifischen methodischen Probleme sich zwischen den Bereichen zum Teil deutlich unterscheiden und dass (3) Zeit, Ort und Sprache der Publikation sowie in geringerem Maße auch die Fallzahl in den vorliegenden Studien mit der methodischen Qualität assoziiert sind.

Bei der Interpretation der Daten muß berücksichtigt werden, dass in den Bereichen Akupunktur und Phytotherapie keine zufällige Stichprobe von Studien gezogen wurde, sondern alle Studien zu zwei Fragestellungen herangezogen wurden. Vor allem bei der Phytotherapie ist davon auszugehen, dass die Ergebnisse zu Hypericum und Echinacea nicht ohne weiteres auf andere Bereiche zu übertragen sind, da unterschiedliche Hersteller unterschiedlich forschungsaktiv bzw. erfahren sind und die Probleme (z.B. bzgl. der Verblindung) bei einzelnen Pflanzenextrakten erheblich variieren. Bei der Akupunktur erscheint nach der Erfahrung des Autors wahrscheinlicher, dass die beobachteten Probleme auch für die meisten anderen Indikationen zutreffen. Bzgl. der Homöopathie ist die Situation anders, da hier alle Studien bis 1996 berücksichtigt wurden.

Die Verwendung von Scores zur „Quantifizierung“ der methodischen Qualität ist in der Fachliteratur umstritten. Während sie von manchen Autoren für die pragmatische Gesamtbeurteilung als sinnvoll angesehen werden (329;352;469), halten andere sie für „sinnlos“ und „potentiell irreführend“ und empfehlen ausschließlich einzelne Qualitätsaspekte separat zu beurteilen (152;225). Die in 3.3.1. präsentierten Analysen nutzen daher beide Vorgehensweisen. Unabhängig davon, ob nun Scores oder Einzelkriterien beurteilt werden, bleibt als Hauptproblem, dass alle formalisierten Qualitätsbeurteilungsmethoden ungenau sind und primär Beurteilungen der Publikationsqualität anstelle der Studienqualität sind. Einzelne Worte (z.B. die Erwähnung, dass die Verblindung der Gruppenzuteilung mithilfe von konsekutiv nummerierten Medikationsbehältern erfolgte) entscheiden darüber, ob eine Studie „gut“ oder „schlecht“ ist. Eine bessere Berichterstattung erfolgt häufiger bei besseren Studien, im Einzelfall kann aber eine formalisierte Qualitätsbeurteilung völlig irreführend sein.

Die häufig unbefriedigende Qualität der randomisierten Studien zu Homöopathie, Akupunktur und Phytotherapie wurde bereits in zahlreichen systematischen Übersichtsarbeiten bemängelt (siehe Kapitel 4). Die Unterschiede zwischen den Therapien bzgl. einzelner Qualitätskriterien verdeutlichen, dass klinische Studien zu verschiedenen komplementärmedizinischen Verfahren durchaus mit unterschiedlichen Schwierigkeiten konfrontiert sind. Verblindungsprobleme liegen v.a. bei der Akupunktur auf der Hand. In Arzneimittelstudien ist es auch einfacher, ein angemessenes Concealment zu gewährleisten. Die einfachste Methode in Studien zu nichtmedikamentösen Interventionen – ein Zettel mit der Zuteilung in einem verschlossenen Umschlag, der nach Aufnahme des Patienten in die Studie geöffnet wird – gilt als nicht absolut sicher. Andere beobachtete Unterschiede sind weniger leicht zu erklären, z.B. warum Homöopathiestudien weniger oft explizit randomisiert sind und Abbrüche und Ausschlüsse weniger gut beschreiben. Das fast vollkommene Fehlen von intent-to-treat Analysen bei praktisch allen älteren, aber auch bei vielen neueren Studien scheint – ähnlich wie bei den Studien zur Migräneprophylaxe mit Propranolol, Metoprolol und Flunarizin (siehe Kapitel 2) – ein bereichsübergreifendes Problem zu sein.

Ursprünglich war geplant gewesen, die Qualitätsbewertungen der eingeschlossenen Studien mit denen zweier ähnlicher Analysen von Moher et al. aus den Jahren 1996 und 1998 zu „schulmedizinischen“ Interventionen zu vergleichen (328;331). Für die Studie von 1996 wurden 229 zwischen 1989 und 1994 publizierte randomisierte Studien aus 7 führenden englischsprachigen Zeitschriften und 6 anderssprachigen Zeitschriften mit dem Ziel ausgewertet, Zusammenhänge zwischen Qualität und Sprache zu untersuchen. Für die Studie von 1998 benutzten Moher et al. 127 randomisierte Studien aus 11 Meta-Analysen zu verschiedenen Fragestellungen, um den Einfluss der Studienqualität auf die Ergebnisse zu untersuchen. Alle Studien wurden mithilfe des Jadadscores bewertet; zusätzlich wurde die Angemessenheit der

Verblindung der Gruppenzuteilung (allocation concealment) beurteilt. Der Autor dieser Schrift bewertete in der 1996er-Analyse die Studien aus der *Deutschen medizinischen Wochenschrift*. Die Verwendung des Jadadscores erfolgte nach identischen Instruktionen. Leider war es nicht möglich, Daten zu den Charakteristika aller von Moher et al. einbezogenen Studien zu bekommen. Dementsprechend kann ein Vergleich nur aufgrund der publizierten Ergebnisse erfolgen. Die mittleren Jadadscores der komplementärmedizinischen und „schulmedizinischen“ Studien waren sehr ähnlich (2,55 und 2,74 bei Moher et al. im Vergleich zu 2,61 bei den komplementärmedizinischen Studien). Die komplementärmedizinischen Studien waren öfter doppelblind, berichteten öfter eine angemessene Methode zur Verblindung der Gruppenzuteilung, beschrieben aber Abbrüche und Ausschlüsse weniger ausführlich. Ohne genauere Kenntnis der „schulmedizinischen“ Interventionen sind diese Daten schwer zu interpretieren. Sie suggerieren jedoch, dass - zumindest in Bezug auf die erfaßten Kriterien - komplementärmedizinische Studien nicht schlechter sind als viele „schulmedizinische.“

Zusammenfassend ist festzustellen, dass es in den Studien zu Akupunktur, Phytotherapie und Homöopathie erheblichen Verbesserungsbedarf gibt. Die vorliegenden Ergebnisse sprechen dafür, dass insbesondere die Beschreibung von Studienabbrüchen und –ausschlüssen sowie deren Berücksichtigung in der statistischen Auswertung unbefriedigend ist. Bei der Publikation von Studien sollten die Autoren – wie auch im Fall der Migräneprophylaxe mit Medikamenten – unbedingt die CONSORT-Leitlinien befolgen (34;333).

3.4.2 Einfluss von Aspekten der methodischen Qualität auf das Studienergebnis am Beispiel der placebokontrollierten Studien zur Homöopathie

Die Untersuchungen zum Einfluss der methodischen Qualität auf das Ergebnis machen deutlich, dass placebokontrollierte Studien zur Homöopathie, die definierte Qualitätskriterien erfüllen, im Mittel geringere Effekte zeigen als Studien, die diese Kriterien nicht erfüllen. Die plausibelste Erklärung hierfür ist, dass in qualitativ weniger hochwertigen Studien mehr Verzerrung (Bias) vorhanden ist und die Wirksamkeit häufig überschätzt wird.

Unsere Ergebnisse sind grundsätzlich ähnlich wie die entsprechender Untersuchungen an „schulmedizinischen“ Studien. Der Einfluß der einzelnen Qualitätsaspekte variiert jedoch zum Teil erheblich. In den 250 von Schulz et al. (406) analysierten Studien aus 33 systematischen Übersichtsarbeiten aus der Cochrane Pregnancy and Childbirth Datenbank erwies sich die Verblindung der Randomisation als stärkster Einflussfaktor (d.h. Studien, die dieses Kriterium erfüllten, hatten deutlich pessimistischere Ergebnisse). Auch der Einfluß von Doppelblindbedingungen war signifikant, eine explizit randomisierte Zuteilung hatte jedoch keinen Einfluss. Wie in unserer Analyse hatten Studien mit vollständigem Follow-up tendenziell

optimistischere Ergebnisse als Studien, bei denen dies nicht der Fall war. Moher et al. (331) analysierten, wie im vorgehenden Abschnitt beschrieben, 127 Studien aus 11 Meta-Analysen zu verschiedenen Interventionen. Die Ergebnisse waren ähnlich wie die von Schulz et al. bezüglich Verblindung der Zuteilung, doppelblinde Studien zeigten jedoch nicht signifikant pessimistischere Resultate. Bei den vorliegenden Studien zur Homöopathie erwiesen sich Doppelblindbedingungen als wichtigster Faktor. Eine mögliche Erklärung hierfür wäre, dass Studien zur Homöopathie häufig bei Studien mit leichteren oder chronischen Erkrankungen durchgeführt werden, bei denen die Beurteilung meist anhand „weicher“ subjektiver Parameter erfolgt, die unter „offenen“ Prüfbedingungen leicht verfälscht werden können.

Die klaren Hinweise auf das Vorliegen von Verzerrungen schwächen die Aussagekraft der Ergebnisse der von uns durchgeführten Meta-Analyse zur Homöopathie (278). Wir stellten unsere Daten außerdem für eine unabhängige Re-Analyse zur Verfügung, bei der der Regressions-Test auf Small Study Bias von Egger et al. zum Einsatz kam (440). Auch diese Untersuchungen deuten darauf hin, dass bessere Studien weniger optimistische Ergebnisse zeigen. Die Ergebnisse von Egger et al. geben auch eine mögliche Erklärung, wie ein so deutlicher Effekt in einer Meta-Analyse auftreten kann, obwohl in Wirklichkeit kein Effekt über Placebo hinaus vorhanden ist. Allerdings weisen die Autoren darauf hin, dass dies kein Beweis dafür ist, dass die positiven Ergebnisse tatsächlich alle durch Bias zustandekamen. Man kann jedoch mit an Sicherheit grenzender Wahrscheinlichkeit davon ausgehen, dass die in unserer Meta-Analyse berechnete Effektgröße zu optimistisch ist und den wahren Effekt über Placebo überschätzt.

In unseren Analysen führte die Anwendung der vordefinierten Schwellenscores, um „bessere“ und „schlechtere“ Studien durch die beiden Qualitätsskalen zu trennen, zu konservativeren Effektschätzern als die Anwendung einzelner Qualitätskriterien. Auch die kumulativen Meta-Analysen deuten darauf hin, dass in unserem speziellen Fall die Qualitätsscores sensible Instrumente waren. Ein eindeutiger linearer Zusammenhang zwischen den Scorewerten und den Ergebnissen existiert jedoch nicht. Hierfür gibt es verschiedene Erklärungen: 1) Sollte die Ausgangshypothese der Meta-Analyse, „alle homöopathischen Interventionen sind Placebos,“ falsch sein, muss unsere Studiensammlung als extrem heterogen angesehen werden, da nicht zu erwarten ist, dass alle homöopathischen Interventionen bei verschiedensten Indikationen immer gleich wirksam sind. Bei einer solchen Heterogenität wäre zu erwarten, dass eine leichte bis mittlere Korrelation im „Rauschen“ nicht mehr zu identifizieren ist. 2) Eine zweite Erklärung wäre, dass die Zusammenhänge zwischen Qualität und Ergebnis eher gering sind. 3) Schließlich ist es möglich, dass die verwendeten Qualitätsscores nicht dazu geeignet sind, einen tatsächlich bestehenden Zusammenhang zu erfassen. In einer Gruppe von 7 Meta-Analysen konnten Emerson et al. ebenfalls keinen klaren Zusammenhang zwischen Qualitätsscores und Ergebnissen feststellen (103). Dies spräche für die Erklärungen 2) oder 3).

Wie bereits erwähnt, ist die Verwendung von Scores zur Beurteilung der Studienqualität unter Methodikern umstritten. Stattdessen wird empfohlen, den Einfluß einzelner Qualitätsaspekte auf das Ergebnis zu prüfen (152). Diese Empfehlung ist grundsätzlich einleuchtend. Unsere Ergebnisse deuten darauf hin, dass Analysen mit Qualitätsscores zu ähnlichen Ergebnissen wie Analysen mit einzelnen Qualitätsaspekten kommen können. Aufgrund ihrer größeren Transparenz und dem konzeptionell logischeren Ansatz erscheinen uns jedoch multivariate Meta-Regressions-Analysen einzelner Qualitätsaspekte als Methode der ersten Wahl.

Einschränkend ist anzumerken, dass in Meta-Analysen im Schnitt nur etwa 10 (439) und dann häufig auch noch heterogene Primärstudien eingehen. In diesen häufigen Fällen ist jedwede empirische Analyse zum Einfluß der Qualität auf die Studienergebnisse schwierig und von begrenzter Aussagekraft. Außerdem wurde bereits darauf hingewiesen, dass formale Qualitätsbeurteilungen – ob sie nun mit Scores oder über Einzelaspekte erfolgen – in aller Regel äusserst grob sind.

Angesichts der eindeutigen Indizien für einen Zusammenhang zwischen Studienqualität und Ergebnissen einerseits und den noch nicht ausgereiften Methoden v.a. zur Qualitätsbeurteilung andererseits besteht ein dringender Bedarf an weiterer empirisch-methodischer Forschung. In der momentanen Situation erscheint es sinnvoll, sowohl Scores wie auch Einzelkriterien bei der Untersuchung auf Zusammenhänge von Qualität und Ergebnis zu verwenden. Alle Interpretationen müssen aber mit der notwendigen Zurückhaltung erfolgen.

3.4.3 Einfluss von Qualitätsmerkmalen auf Therapieerfolgswerten in Studien unterschiedlicher Designs am Beispiel Akupunktur bei chronischen Kopfschmerzen

Die in 3.3.3. präsentierten Ergebnisse zeigen, dass im untersuchten Beispiel Akupunktur bei chronischen Kopfschmerzen 1) deutliche Unterschiede zwischen randomisierten und nichtrandomisierten Studien in Bezug auf Patientenmerkmale, Akupunkturbehandlungen und Therapieerfolgswerten bestehen; 2) randomisierte Studien im Schnitt mehr Kriterien zur Angemessenheit der Beobachtung und Berichterstattung erfüllen als nichtrandomisierte Studien, dass aber ein relevanter Teil der randomisierten Studien sehr wenige Kriterien erfüllt und einige prospektiv-unkontrollierte Studien gute Qualität haben; 3) die verfügbaren nichtrandomisierten Studien nur sehr wenig Zusatzinformationen bzgl. prognostischer Faktoren, Langzeiterfolg, Nebenwirkungen und Generalisierbarkeit erbringen; und 4) ein besseres Studiendesign (randomisiert vs. nichtrandomisiert) und bessere Beobachtung und Berichterstattung (unabhängig vom Design) mit geringeren Therapieerfolgswerten assoziiert sind.

Bei der Interpretation der Daten sind wiederum einige Probleme zu beachten. Während die randomisierten Studien von mindestens zwei Reviewern bewertet wurden, erfolgte dies für die

nichtrandomisierten Studien nur durch den Autor, allerdings zweimal im Abstand von mehreren Monaten. Daher kann nicht sicher ausgeschlossen werden, dass bei einzelnen Studien Extraktionsfehler gemacht wurden. Die Kriterienliste zur Beurteilung der Angemessenheit von Beobachtung und Berichterstattung ist nicht formal validiert. Sie kann daher lediglich als Versuch angesehen werden, die Bewertung von Qualitätsaspekten jenseits gängiger methodischer Kriterien zu standardisieren. Außerdem erfolgten multiple statistische Testungen ohne Anpassung der p-Werte. Signifikante Ergebnisse müssen daher zurückhaltend interpretiert werden.

In der empirisch-methodischen Literatur gibt es mehrere Untersuchungen zu Unterschieden in den Ergebnissen randomisierter und nichtrandomisierter, kontrollierter Studien (37;72;199;256;291). Die Ergebnisse dieser Studien sind uneinheitlich, deuten aber darauf hin, dass gute nichtrandomisierte, kontrollierte Studien ähnliche Effekte zeigen wie randomisierte, dass jedoch bei Diskrepanzen nichtrandomisierte Studien etwas häufiger größere Effekte zeigen. Wie bereits erwähnt, gibt es gute Evidenz, dass bei randomisierten Studien qualitativ hochwertigere Methoden mit negativeren Ergebnissen assoziiert sind. Nach Kenntnis des Autors gibt es bisher jedoch keine Analysen, die Qualität und Ergebnisse (vorher-nachher Unterschiede) von randomisierten und nichtrandomisierten Studien einschließlich unkontrollierter Untersuchungen vergleichen.

Die geringeren Therapieerfolgsraten in randomisierten Studien einerseits und grundsätzlich in besseren Studien unabhängig vom Design andererseits suggerieren, dass in schlechten Studien der Therapieerfolg häufig überschätzt wird. Möglicherweise nehmen an randomisierten Studien Patienten mit schlechterer Prognose teil. Man könnte auch spekulieren, dass Akupunktur unter experimentellen Bedingungen bzw. bei den Unsicherheiten in einer randomisierten Studie weniger gut wirkt.

Aus Sicht des Autors werden bei der formalen Bewertung der Qualität klinischer Studien methodische Kriterien überbetont, während klinische Qualitätsaspekte gar nicht oder ausschließlich unsystematisch bewertet werden. Dies ist bis zu einem gewissen Grad verständlich, da methodische Kriterien weitgehend universell gelten, während klinische Qualitätsaspekte stark kontextabhängig sind. Dennoch macht es aber wenig Sinn, Qualität bei Erkrankungen wie Migräne ausschliesslich über die Nennung von Randomisation, Doppelblindbedingungen und die Beschreibung von Studienabbruchern zu bestimmen. 10 der 24 randomisierten Akupunkturstudien erfüllten weniger als 5 der – äußerst basalen – Kriterien zur Angemessenheit der Beobachtung und Berichterstattung. Das macht deutlich, dass ein relevanter Teil der beurteilten randomisierten Studien kaum klinisch sinnvoll interpretierbar ist.

Die Schlussfolgerung des von unserer Autorengruppe durchgeführten systematischen Reviews der randomisierten Studien zur Akupunktur bei chronischen Kopfschmerzen (321) war, dass die vorhandene Evidenz für eine Effektivität dieser Therapie bei der Prophylaxe von Migräneattacken

und Spannungskopfschmerzen spricht, dass jedoch zusätzliche, bessere und größere Studien notwendig sind, um diese positiven Befunde zu sichern und die Größe des Effektes und die Generalisierbarkeit zu beurteilen. Die nichtrandomisierten Studien verstärken den Eindruck der Effektivität. Die Menge der durch den Einbezug dieser Studien gewonnenen Zusatzinformation ist aber im Verhältnis zum Aufwand enttäuschend. Die Unterschiede zwischen randomisierten und nichtrandomisierten Studien legen die Vermutung nahe, dass an randomisierten Studien nicht die „typischen“ Akupunkturpatienten teilnehmen. Da die nichtrandomisierten Studien aber ebenfalls – soweit nachvollziehbar – ausnahmslos unizentrisch durchgeführt waren und die Patientenrekrutierung kaum angemessen beschrieben war, ist deren Generalisierbarkeit ähnlich unklar wie im Fall der randomisierten Studien. Auch bezüglich der sonstigen, aus nichtrandomisierten Studien gut zu gewinnenden Informationen war der Erkenntniszugewinn gering. Die Schlussfolgerungen unseres systematischen Reviews zu Akupunktur bei chronischen Kopfschmerzen hätten sich daher nur in einem Punkt deutlich geändert: Neben besseren randomisierten Studien sind auch bessere nichtrandomisierte Studien notwendig.

4 Analyse der vorliegenden systematischen Übersichtsarbeiten zu Akupunktur, Phytotherapie und Homöopathie

4.1 Hintergrund

Mit der zunehmenden Zahl systematischer Übersichtsarbeiten entsteht die Möglichkeit, eine weitere Metaebene einzuführen: Analysen von systematischen Übersichtsarbeiten bzw. Meta-Meta-Analysen (291). Diese können verschiedenen Zielen dienen:

1. Sie können einen groben Überblick über die Evidenzlage in größeren Themenbereichen geben. Aufgrund der erheblichen Entfernung vom Originalforschungsgegenstand besteht hier aber eine sehr große Gefahr der Übervereinfachung.
2. Sie erleichtern Ärzten, Patienten und Entscheidungsträgern den Zugang und die Orientierung zu systematischen Übersichtsarbeiten in Bereichen, über die sie sich mit geringem Zeitaufwand informieren wollen.
3. Sie erlauben die Beschreibung von Vorgehensweisen, Merkmalen und Qualität im Sinne einer „Epidemiologie“ von systematischen Übersichtsarbeiten. Dadurch lassen sich insbesondere Schwachstellen identifizieren, die in zukünftigen Reviews vermieden werden sollten.
4. Wenn mehrere systematische Übersichtsarbeiten zu einer Fragestellung vorliegen, ist es möglich zu überprüfen, ob diese sich im Hinblick auf Methoden und Ergebnisse unterscheiden. Ist dies der Fall, können die Gründe für Diskrepanzen untersucht werden.

Die komplementärmedizinischen Verfahren Akupunktur, Phytotherapie und Homöopathie sind aufgrund ihrer großen Verbreitung bei gleichzeitiger Umstrittenheit Bereiche, zu denen sich eine große Zahl von Personen eine Meinung bilden möchte, ohne sich im Detail in die Literatur einarbeiten zu wollen. Im Rahmen des Complementary Medicine Field der Cochrane Collaboration - einem weltweiten Netzwerk von Patienten, Ärzten und Wissenschaftlern mit dem Ziel systematische Reviews in allen Bereichen der Medizin zu erstellen, zu aktualisieren und zugänglich zu machen (2) – wurde unter Federführung des Autors eine Sammlung und Analyse der verfügbaren systematischen Übersichtsarbeiten mit den obengenannten Zielen (mit besonderer Betonung der bei Ziel 1 genannten Einschränkung) durchgeführt. Die Ergebnisse werden in drei Abschnitte unterteilt:

1. Die Zusammenstellung und -fassung der vorliegenden systematischen Übersichtsarbeiten zu Akupunktur, Phytotherapie und Homöopathie (dieser Teil der Arbeit ist bereits publiziert (279;288;289));
2. Die Beschreibung von inhaltlichen und methodischen Merkmalen sowie der Qualität;
3. Die primär qualitative, vergleichende Analyse der Übersichtsarbeiten zu gleichen Fragestellungen.

4.2 Methodik

4.2.1 Ein- und Ausschlusskriterien

Um eingeschlossen zu werden, musste eine Übersichtsarbeit folgende Kriterien erfüllen:

1. In der Arbeit mussten prospektive (nicht notwendigerweise kontrollierte) klinische Studien zu Akupunktur (als Akupunktur bezeichnete Techniken mit Nadelinsertion sowie verwandte Techniken mit Stimulation explizit so bezeichneter „Akupunkturpunkte“ ohne Nadelinsertion wie z.B. Akupressur oder Laserakupunktur), Phytotherapie (Applikation von Extrakten oder anderen Zubereitungen von Pflanzen; isolierte oder synthetisch hergestellte, in Extrakten vorkommende Einzelsubstanzen wurden nicht berücksichtigt) oder Homöopathie (Applikation homöopathischer Arzneimittel) am Menschen referiert werden.
2. Mindestens eines der folgenden methodischen Kriterien musste erfüllt sein:
 - a) explizite Beschreibung der Literatursuchmethoden und explizite Beschreibung von Ein- und Ausschlusskriterien;
 - b) explizite Beschreibung einer formalen Methode zur Beurteilung der Qualität der Primärstudien (z.B. mithilfe eines Scores);
 - c) Durchführung einer quantitativen Meta-Analyse mit expliziter Beschreibung der statistischen Methoden.
3. Verfügbarkeit als Publikation oder schriftlicher Bericht in Zeitschriften, Büchern, Dissertationen oder im Internet. Vor 1989 publizierte Reviews und solche, die lediglich als Abstract verfügbar waren, wurden nicht berücksichtigt.
4. Der Schwerpunkt der bearbeiteten Fragestellung des Reviews bezog sich auf die Wirksamkeit der Intervention (nicht Schwerpunkt auf Diagnose, Nebenwirkungen, Risikofaktoren etc.).

Eingrenzungen bzgl. der Publikationssprache gab es keine. Krankheitsbezogene systematische Übersichtsarbeiten zu mehreren Interventionen (z.B. zu allen Therapien bei Postzosterneuralgie) wurden nur eingeschlossen, wenn mindestens vier Primärstudien zu einer der drei komplementären Therapien bearbeitet worden waren (Begründung: Nur wenn eine gewisse Anzahl an Primärstudien zu einer bestimmten Therapieform eingeschlossen wird, ist wahrscheinlich, dass diese in Abstracts oder Schlüsselworten überhaupt genannt wird und die Übersichtsarbeit so mit Hilfe gängiger Datenbanksuchen zu identifizieren ist).

4.2.2 Literatursuche

Primäre Quelle zur Identifikation potentiell relevanter Reviews war die Datenbank des Cochrane Complementary Medicine Field. Dieses ursprünglich aus der komplementärmedizinischen Datenbank CISCOM (Research Council of Complementary Medicine, London) hervorgegangene Register (verfügbar unter www.compmed.umm.edu) enthält neben Referenzen zu kontrollierten klinischen Studien zu unkonventionellen medizinischen Verfahren Referenzen zu systematischen Übersichtsarbeiten bzw. zu Referenzen, die möglicherweise systematische Übersichtsarbeiten sein könnten. Diese wurden aus einer Vielzahl von Datenbanken (z.B. Medline, Embase, AMED, Current Contents, Datenbank des Zentrums für naturheilkundliche Forschung München) mit systematischen Methoden gesammelt. Zusätzlich wurden in Medline (Silver Platter 1989 bis Juli 2000 mithilfe einer Standardsuchstrategie für systematische Übersichtsarbeiten (149) kombiniert mit Suchbegriffen für die Therapie: acupuncture or acupressure; medicinal plant exploded sowie Namen 50 medizinisch häufig verwendeter Pflanzen; homeopath* oder homoeopath*) und in der Cochrane Library (zuletzt Ausgabe 2000, 3) Suchen durchgeführt. Außerdem wurden die Literaturlisten eingeschlossener Arbeiten und von mehreren Lehrbüchern zum Thema nach potentiell relevanten Referenzen gescreent.

4.2.3 Selektionsprozess

Alle in der Datenbank des Cochrane Complementary Medicine Field gefundenen, potentiell relevanten Referenzen wurden in einem ersten Schritt von 2 Reviewern, die beide umfangreiche Erfahrung in der Durchführung von systematischen Reviews und bzgl. komplementärmedizinischer Datenbanken hatten, unabhängig durchgesehen. Dabei wurden diejenigen aussortiert, bei denen sich beide Reviewer sicher waren (meist aufgrund der Kenntnis des jeweiligen Artikels), dass dies Arbeiten waren, die die Einschlusskriterien mit Sicherheit nicht erfüllten. Die Abstracts der in Medline und der Cochrane Library gefundenen Zitate wurden durch einen Reviewer durchgesehen. Kopien aller potentiell relevanten Arbeiten wurden besorgt und von einem oder zwei bis drei Reviewern mit Hilfe einer Checkliste auf die Erfüllung der Einschlusskriterien geprüft. Aufgrund begrenzter Ressourcen war es nicht möglich, alle potenziell relevanten Arbeiten von zwei oder mehr Reviewern prüfen bzw. extrahieren zu lassen. Insgesamt 115 Übersichtsarbeiten erfüllten die Einschlusskriterien.

4.2.4 Extraktion und Beurteilung

Ein (für 46% aller eingeschlossenen Übersichtsarbeiten) oder zwei bzw. drei (54%) Reviewer extrahierten mithilfe eines in einem Vorlauf getesteten Formulars bibliographische Angaben, Informationen zu Fragestellung, Interventionen, Indikationen, Einschlusskriterien, Suchmethoden, Extraktion und Qualitätsbeurteilung der Primärstudien, Anzahl und Art der einbezogenen Primärstudien sowie zu Zahl der in diesen untersuchten Patienten, zu Methoden der Ergebniszusammenfassung, Schlussfolgerungen und Angaben zum Sponsoring des Reviews. Die Daten wurden in eine SPSS-Datei eingegeben und deskriptiv statistisch ausgewertet.

Für die tabellarische Zusammenfassung der Reviews wurden Ergebnisse und Schlussfolgerungen soweit möglich in den Worten der jeweiligen Autoren herangezogen. Zur Orientierung wurden folgende methodische Merkmale kategorisiert: Ob eine ausreichende Literatursuche erfolgte (wenn auch in anderen Quellen als in Medline bzw. einer Datenbank gesucht wurde), ob die Einschlusskriterien ausreichend beschrieben worden waren, ob die Qualität der Primärstudien mit expliziten Methoden beurteilt, für jede Primärstudie eine Zusammenfassung des Ergebnisses (über einen vote count hinaus) präsentiert und ob eine quantitative Meta-Analyse durchgeführt worden war.

Darüber hinaus wurde die Qualität der eingeschlossenen Reviews mithilfe der Skala von Oxman et al. (355;356) beurteilt. Diese umfasst folgende 10 Fragen:

1. Sind die Literatursuchmethoden beschrieben?
2. War die Literatursuche umfassend?
3. Wurden die Einschlusskriterien für die Primärstudien beschrieben?
4. Wurden Verzerrungen bei der Studienselektion vermieden?
5. Wurden Kriterien zur Beurteilung der Qualität der Primärstudien beschrieben?
6. Wurde die Qualität der Primärstudien angemessen beurteilt?
7. Wurden die meta-analytischen Methoden beschrieben?
8. Waren die meta-analytischen Methoden angemessen?
9. Sind die Schlussfolgerungen aufgrund der präsentierten Ergebnisse nachvollziehbar?
10. Wie ist die Gesamtqualität der Übersichtsarbeit?

Die Fragen 1 bis 9 werden mit ja (1 Punkt), teilweise (0,5 Punkte) oder nein (0 Punkte) beantwortet, bei Frage 10 ist eine Skala von 1 (schwerwiegende Schwächen) bis 7 (keine Schwächen) vorgegeben.

4.2.5 Vergleich der Übersichtsarbeiten zu gleichen Fragestellungen

Ziel dieses Vergleichs war zu untersuchen (1) wie häufig systematische Übersichtsarbeiten zu gleichen Fragestellungen Unterschiede bzgl. Methoden, Ergebnissen und Schlussfolgerungen aufweisen und (2) was die Gründe für etwaige Nichtübereinstimmungen sind. Dabei wurden die Einschlusskriterien etwas enger gefasst, um zu gewährleisten, dass nur systematische Reviews mit einer ausreichenden Qualität berücksichtigt wurden. Jedes der folgenden Kriterien musste erfüllt sein: 1. Die Literatursuche war beschrieben. 2. Basale Einschlusskriterien waren definiert. 3. Es war eine formale Beurteilung der methodischen Qualität der Primärstudien mithilfe eines Scoresystems oder über Einzelmerkmale erfolgt oder es waren methodische Merkmale der Primärstudien transparent beschrieben. 4. Wenn Effektstärkeberechnungen durchgeführt worden waren, mußten die entsprechenden Methoden beschrieben sein. Wenn keine Effektstärkeberechnungen erfolgt waren, mußte die jeweilige Methode der Ergebniszusammenfassung beschrieben sein oder es mußte für jede Studie eine deskriptive Zusammenfassung der Hauptergebnisse präsentiert sein. Die Überprüfung der Einschlusskriterien erfolgte durch den Autor.

Die Fragestellung wurde als gleich angesehen, wenn 1. die Interventionen gleich waren (z.B. Hypericumextrakte), 2. die bearbeitete Indikation identisch war (z.B. Depression), 3. die Kontrollinterventionen identisch waren (z.B. Placebo und andere Medikamente) und 4. gleiche oder ähnliche Zielkriterien (z.B. Depressionsscores) verwendet wurden. Außerdem durfte die Differenz der Publikationsjahre maximal 3 betragen (z.B. 1997-2000). Wenn ein Review zwei oder mehr Indikationen (z.B. Migräne und Spannungskopfschmerzen) bearbeitete, wurde er nur berücksichtigt, wenn die einzelnen Indikationen im Review klar trennbar waren (d.h. wenn erkennbar war, welche Ergebnisse auf Migräne und welche auf Spannungskopfschmerzen bezogen waren).

Der Selektionsprozess erfolgte in zwei Stufen. In einem ersten Schritt wurden alle gefundenen Übersichtsarbeiten zu Akupunktur, Phytotherapie und Homöopathie, die den in 4.2.1. genannten Einschlusskriterien entsprachen, geprüft und grob nach Fragestellungen klassifiziert. Alle Übersichtsarbeiten, die *ähnliche* Fragestellungen bearbeiteten, wurden detailliert vergleichend ausgewertet (siehe unten). Im zweiten Schritt wurde aus den hier bearbeiteten Reviews diejenigen selektiert, die den enger gefassten Einschlusskriterien für ausreichend hochwertige systematische Reviews entsprachen, von unabhängigen Reviewergruppen erarbeitet worden waren und die entsprechend den oben genannten Bedingungen *gleiche* Fragen bearbeiteten.

Aus allen Reviews zu *ähnlichen* Fragestellungen wurden Informationen zu folgenden Punkten in eine Tabellenmatrix übertragen: Details zu Literatursuche (welche Datenbanken, andere Suchmethoden), zu Einschlusskriterien (bzgl. Patienten, Test- und Kontrollinterventionen, Zielkriterien, Studiendesign, Sprachen, sonstige), zum Selektionsprozess (ob beschrieben oder

nicht, Anzahl der Studien auf den einzelnen Selektionsniveaus), zur Datenextraktion, zu den benutzten Qualitätsbeurteilungs- und Ergebniszusammenfassungsmethoden, die Anzahl der eingeschlossenen Primärstudien (unter Berücksichtigung des Publikationsjahres), Ergebnisse, Statements zur methodischen Qualität der Primärstudien und zu Schlussfolgerungen. Für jede Fragestellung wurde eine Liste aller Primärstudien erstellt, die in mindestens einem Review eingeschlossen waren, um den Einfluss des Alters des Reviews auf die Ergebnisse und Schlussfolgerungen beurteilen zu können. Hauptzielkriterium war die Zahl der jeweils eingeschlossenen Primärstudien bzw. der Grad der Übereinstimmung unter Berücksichtigung des Publikationsjahres. Unterschiede von mehr als 25% wurden als relevant klassifiziert.

4.3 Ergebnisse

4.3.1 Zusammenstellung der systematischen Übersichtsarbeiten zur Akupunktur

Von 48 potentiell relevanten Übersichtsarbeiten zur Akupunktur, die durch den Screeningprozess vorselektiert waren, erfüllten 39 (publiziert in 45 Veröffentlichungen bzw. Berichten) die Einschlusskriterien (siehe Tabellen 17 bis 21). Dabei wurden aktualisierte Versionen eines Reviews von der gleichen Autorengruppe als ein Review gezählt. 5 Reviews wurden ausgeschlossen, da sie nicht als systematisch klassifiziert wurden (Nichterfüllung von Einschlusskriterium 2) (33;51;84;105;387) und weitere 4, weil nicht die Beurteilung der Wirksamkeits sondern methodische Aspekte im Mittelpunkt standen (19;161;473;475).

Ein relevantes Problem in der Akupunkturforschung ist die Frage, inwieweit die Nadelung bzw. Stimulation definierter Punkte spezifische Effekte hat. In den vorliegenden Primärstudien wurde eine große Zahl unterschiedlicher „Placebo“-Techniken angewendet. Gemeinsam ist vielen, dass sie nicht als physiologisch inert angesehen werden können (19). Die Bezeichnung Placeboakupunktur für derartige Techniken ist daher problematisch und wird im folgenden vermieden. Stattdessen wird die Bezeichnung Sham oder Scheinakupunktur benutzt.

Wie zu erwarten, beschäftigte sich eine relevante Anzahl von Reviews mit dem Thema Schmerz (siehe Tabellen 17 und 18). 3 Reviews (publiziert 1989 (361;457), 1990 (455) und 2000 (131)) bearbeiteten chronischen Schmerz allgemein. Die Studiensammlungen – insgesamt wurden 72 Primärstudien in den drei Übersichtsarbeiten eingeschlossen – und die Reviewmethoden unterschieden sich erheblich, die vorhandene Evidenz wurde aber in allen Fällen als „inconclusive“, also nicht für klare Schlussfolgerungen ausreichend, beurteilt. Rücken- bzw. Nackenschmerzen waren das Thema von insgesamt 6 Reviews (127;293;415;449;465;466;502). Die Gesamtzahl der bearbeiteten Primärstudien war jeweils 11 für Rücken- und Nackenschmerz. Die Schlussfolgerungen bzgl. Rückenschmerzen waren deutlich unterschiedlich (siehe Tabelle 17),

während bzgl. Nackenschmerzen die Evidenz wiederum als „inconclusive“ beurteilt wurde. 6 Übersichten zu insgesamt 25 Primärstudien behandelten Kopfschmerzen (siehe Tabelle 18). Der umfangreichste Review zog eingeschränkt positive Schlussfolgerungen (321), während in den übrigen, kleineren Übersichten die Evidenz abermals als „inconclusive“ gewertet wurde (151;312;450;451;470). Die Ergebnisse der vorliegenden Studien zu akuten Zahnschmerzen und zur temporomandibulären Dysfunktion wurden in zwei weiteren Reviews als vielversprechend beurteilt (120;128).

Tabelle 17:
Systematische Übersichtsarbeiten zu klinischen Studien zur Akupunktur bei Schmerzzuständen – Teil I

Autor & Jahr	Indikation	Kontrolle	Studien	Method 1-2-3-4-5	Ergebnisse	Schlussfolgerungen
Chronische Schmerzen						
Ezzo 2000 (131)	Chronische Schmerzen	Sham, Placebo, keine Beh., Standard	51 RCT	j-j-j-j-n	Positive Ergebnisse in 21 Studien, negative in 3, neutrale in 27. Bessere Studien öfter negativ oder neutral	Hinweise, dass Akup. einer Nichtbehandlung überlegen ist, Evidenz unklar im Vergleich zu Sham, Placebo und Standard
ter Riet 90/89 (455;457)	Chronische Schmerzen	Sham, keine Behandlung, Standard	51 CCT	j-j-j-j-n	Studien klein und von meist schlechter Qualität. 24 mit positiven, 27 mit negativen Ergebnissen. Bessere Studien öfter neg.	Keine überzeugenden Belege für die Wirksamkeit der Akupunktur bei chronischen Schmerzen
Patel 89 (361)	Chronische Schmerzen	Sham, keine Behandlung, Standard	14 RCT	n-j-n-j-j	Wahrscheinlichkeit einer Besserung unter Akupunktur im Mittel um 18% ($p < 0.01$) größer	Die vorhandene Evidenz ist positiv, aufgrund von methodischen Schwächen aber nicht überzeugend
Rücken- und Nackenschmerzen						
Smith 2000 (415)	Rücken- & Nackenschmerzen	Sham, keine Behandlung, Standard	13 RCT	j-j-j-j-n	5 Studien positiv, 8 negativ; bessere Studien öfter negativ	Keine überzeugenden Belege für die Wirksamkeit der Akupunktur bei Rücken- und Nackenschmerzen
White 99 (502)	Nackenschmerzen	Sham, keine Behandlung, Standard	14 RCT	j-j-j-j-n	7 Studien positiv, 7 negativ. Von den 8 besseren Studien 5 negativ, 3 positiv	Keine überzeugenden Belege für die Wirksamkeit der Akupunktur bei Rückenschmerzen
van Tulder 99 (465;466)	LWS-Schmerzen	Sham, keine Behandlung, Standard	11 RCT	j-j-j-j-n	Studienautoren interpretieren 8 Studien als positiv, die Reviewautoren 2. Methodische Qualität der Studien überwiegend schlecht	Autoren können keine Empfehlung aussprechen, Akupunktur einzusetzen. Gute, große Studien notwendig
Ernst 98 (127)	Rückenschmerzen	Sham, keine Behandlung, Standard	12 RCT	j-j-j-j-j	OR einer Besserung im Vgl. zu allen Kontrollen 2,30 (95%KI 1,28-4,13), mit Sham 1,37 (0,84-2,25). Mehrheit der Studien von guter Qualität	Akupunktur besser als verschiedene Vergleichsinterventionen, aber keine überzeugenden Belege im Vergleich zu Placebo
Longworth 97 (293)	Rückenschmerzen (mit Ausstrahlung)	unklar	1 RCT, 6 CCT, 31 unkontroll.	t-t-n-j-n	Studien meist von schlechter Qualität; Mehrheit der Patienten scheint zu profitieren	“There may be a role for acupuncture treatment of lumbar disk protrusions and sciatica”
ter Riet 89 (449)	Rücken- & Nackenschmerzen	unklar	16 RCT, 6 CCT	j-t-j-n-n	Studien meist von schlechter Qualität. Nur Ergebnisse einzelner besserer Studien beschrieben	Aufgrund der geringen Qualität der Studien sichere Schlussfolgerungen nicht möglich

Method.: 1 = ausreichende Literatursuche, 2 = explizite Einschlusskriterien, 3 = formale Qualitätsbeurteilung, 4 = Ergebnis aller Einzelstudien zusammengefasst, 5 = Meta-Analyse durchgeführt; j = ja, t = teilweise, n = nein, x = nicht anwendbar, ? = unklar
RCT = randomized controlled trials = randomisierte Studien, CCT = controlled clinical trials = kontrollierte Studien (randomisiert und/oder nichtrandomisiert), KS = Kohortenstudien, UKS = unkontrollierte Studien; OR = Odds ratio, RR = Relatives Risiko, 95%KI = 95%-Konfidenzintervall

*Indikationsorientierter Review zu verschiedenen Interventionen; ** Indikationsorientierter Review zu verschiedenen komplementären Interventionen

Tabelle 18:
Systematische Übersichtsarbeiten zu klinischen Studien zur Akupunktur bei Schmerzzuständen – Teil 2

Autor & Jahr	Indikation	Kontrolle	Studien	Method. 1-2-3-4-5	Ergebnisse	Schlussfolgerungen
Kopfschmerzen						
McCrory 2000* (312)	Spannungskopfschmerz	Sham, Physiotherapie	6 RCT	j-j-j-j-n	3 von 4 Studien vs. Sham positiv (beste negativ), Physiotherapie besser in 1 von 2 Studien	Evidenz nicht ausreichend um Schlussfolgerungen zu ziehen. Weitere Studien notwendig
Melchart 99 (321)	chronische Kopfschmerzen	Sham, keine Behandlung, Standard	22 RCT	j-j-j-j-j	Mehrheit der 14 Studien vs. Sham zumindest mit Trend zugunsten der Akupunktur. Studien vs. andere Interventionen widersprüchlich	Vorliegende Evidenz positiv, aber nicht ausreichend
Goslin 99 (151)	Migräne	Sham, keine Behandlung, Standard	6 RCT	j-j-j-j-n	2 von 3 Studien vs. Sham positiv, ähnliche Effekte wie medikamentöse Prophylaxe in 2 Studien	Evidenz reicht nicht aus, um Schlussfolgerungen zu ziehen
Vernon 99** (470)	Kopfschm. Typ Spannung & zervikogen	Sham, keine Behandlung, Standard	8 RCT	j-j-j-j-n	2 von 4 Studien vs. Sham positiv, Ergebnisse vs. Physiotherapie widersprüchlich	Evidenz reicht nicht aus, um Schlussfolgerungen zu ziehen
ter Riet 89 (450)	Migräne & Spannungskopfschmerz	Sham, andere Behandlung	7 RCT, 1 CCT	j-t-j-n-n	Studien aufgrund von geringer Fallzahl und methodischen Schwächen nicht interpretierbar	Evidenz reicht nicht aus, um Schlussfolgerungen zu ziehen
ter Riet 89 (451)	Gesichtsschmerz	Sham	2 RCT	j-t-j-j-n	Studien von geringer Qualität	Evidenz reicht nicht aus, um Schlussfolgerungen zu ziehen
Verschiedene Schmerzzustände						
Ernst 98 (120)	akute Zahnschmerzen	Sham, keine Behandlung, Standard	11 RCT, 5 CCT	j-t-j-j-n	Mehrheit der Studien mit positiven Ergebnissen	Akupunktur scheint effektiv zu sein; weitere Studien jedoch notwendig
Ernst 99 (128)	temporo-mandibul. Dysfunktion	andere und keine Behandlung	3 RCT	j-j-j-j-n	3 Vergleiche mit Standard und 2 mit Nichtbehandlungen positiv	Akupunktur scheint effektiv zu sein; weitere Studien jedoch notwendig

Abkürzungen siehe Tabelle 17

Insgesamt 5 systematische Übersichtsarbeiten liegen zur Behandlung rheumatischer Erkrankungen vor (siehe Tabelle 19). Die vorliegende, limitierte Evidenz deutet darauf hin, dass Akupunktur bei Fibromyalgie wirksam sein könnte (39). Bei der Behandlung von Osteoarthrosen ergaben sich sowohl unter Verum- als auch unter Sham-Akupunktur Verbesserungen hinsichtlich Schmerz und Funktion, jedoch keine Unterschiede zwischen Verum und Sham (107). Bzgl. weiterer rheumatischer Erkrankungen ist die vorliegende Evidenz nicht ausreichend, um zuverlässige Schlussfolgerungen ziehen zu können (200;265;453).

Bei der Unterstützung der Raucherentwöhnung ist Akupunktur nicht wirksamer als eine Sham-Akupunktur. Diese Schlussfolgerung findet sich konsistent in allen vorliegenden systematischen Reviews zum Thema (266;454;456;503-505), bei denen insgesamt 22 Primärstudien einbezogen worden waren (siehe Tabelle 19). Allerdings zeigten sich auch keine Unterschiede zu belegterweise wirksamen Vergleichsinterventionen (z.B. Nikotinpflaster). Der einzige systematische Review zur unterstützenden Entzugsbehandlung bei Alkohol- und

Heroinabhängigkeit datiert aus dem Jahre 1990 (456); zu dieser Zeit reichte die vorliegende Evidenz nicht für klare Schlussfolgerungen aus. Aktuellere Übersichten zum Thema (z.B. (51)) sind lediglich als narrative Reviews einzustufen.

Die Forschung zur Behandlung von Übelkeit und Erbrechen nach Operationen (siehe Tabelle 20), in der Frühschwangerschaft oder unter Chemotherapie fällt deutlich aus dem Rahmen der übrigen Akupunkturforschung heraus. Zum einen erfolgt die Behandlung fast ausschliesslich an einem einzigen Punkt (Pe 6), zum anderen erfolgt die Stimulation des Punktes in der Mehrzahl der Studien nicht durch Nadelstiche, sondern mithilfe eines Akupressurbandes. Beim postoperativen Erbrechen sprechen die in den Reviews eingeschlossenen insgesamt 24 Studien eindeutig für eine punktspezifische Wirksamkeit (7;214;270;474). Bei Übelkeit in der Frühschwangerschaft ist die Mehrzahl der Studien (insgesamt 8 in den Reviews) ebenfalls positiv, die methodisch beste Studie zeigt jedoch keinen Effekt im Vergleich zur Shambehandlung (7;214;474). Auch bzgl. der Vorbeugung bzw. der Behandlung des chemotherapieinduzierten Erbrechens liegen vielversprechende Ergebnisse vor, jedoch sind die Studien vor der Einführung hochpotenter Antiemetika durchgeführt worden und heute nur noch von begrenzter Relevanz (474).

Die vorliegenden Studien zum Asthma (insgesamt 15 in den vorliegenden Reviews (246;280;290;452)) haben allesamt äußerst geringe Stichprobenumfänge und widersprüchliche Ergebnisse. Insgesamt sprechen die Studien allenfalls für eine sehr begrenzte Effektivität der geprüften Therapiestrategien.

Tabelle 19:
Systematische Übersichtsarbeiten zu klinischen Studien zur Akupunktur bei rheumatischen Erkrankungen und Sucht

Autor Jahr	& Indikation	Kontrolle	Studien	Method. 1-2-3- 4-5	Ergebnisse	Schlussfolgerungen
Rheumatische Erkrankungen						
Berman 99 (39)	Fibromyalgie	Sham, Standard	3 RCT, 3 KS	j-j-j-j-n	Akupunktur besser als Sham bei Global- und Symptombewertung	Vorhandene Evidenz positiv, weitere Studien notwendig
Ernst 97 (107)	Osteoarthrose	Sham, keine Behandlung, Standard	13 RCT	j-t-n-j-n	Sowohl Verum- als auch Sham-Akupunktur mit positivem Effekt, jedoch kein Unterschied zwischen beiden	Überlegenheit über Sham-Akupunktur nicht belegt
Lautenschläger 97 (265)	entzündlich-rheumat. Erkrank.	Sham, keine Behandlung, Standard	2 RCT, 7 CCT, 9 KS	n-t-n-j-n	Widersprüchliche Ergebnisse, viele Studien von geringer Qualität	Akupunktur kann nicht empfohlen werden bei rheum. Arthritis, Spondylarthropathie, Lupus eryth., Sklerodermie
Jacobs 91** (200)	rheumat. Erkrank.	Sham, Standard	23 CCT	t-j-n-j-n	7 Studien positiv, 13 keine Wirkung über Sham oder Standard, 3 Studien unklar	Keine ausreichende Evidenz für eine Therapieempfehlung
ter Riet 89 (453)	rheumatoide Arthritis	Sham	1 RCT, 2 CCT	j-t-j-n-n	Nur eine Studie zusammengefasst: positiver Effekt auf Schmerz, nicht auf Entzündung	Evidenz reicht nicht aus, um Schlussfolgerungen zu ziehen
Sucht						
White 2000/99 (503;505)	Raucherentwöhnung	Sham, keine Behandlung, Standard	20 RCT	j-j-j-j-j	Akupunktur vs. Sham: OR Aufhören 1,22 (95%KI 0,99-1,49) nach Beh. und 1,02 (0,72-1,43) nach 12 Monaten. Kein Unterschied zu anderen Beh., besser als keine Behandlung	Keine Evidenz, dass Akupunktur über eine Sham-Akupunktur hinaus spezifische Effekte bei der Raucherentwöhnung hat
White 97 (504)	Raucherentwöhnung	Sham	7 RCT	j-j-j-j-n	In 6 von 7 Studien keine signifikant unterschiedlichen Erfolgsraten	Keine Evidenz, dass Akupunktur über eine Sham-Akupunktur hinaus spezifische Effekte bei der Raucherentwöhnung hat
Law 95* (266)	Raucherentwöhnung	Sham, keine Behandlung, Standard	8 RCT	t-t-n-n-j	Im Vgl. zur Kontrolle 3% (95%KI -1 bis 6%) mehr Teilnehmer unter Verumakupunktur nicht rückfällig	Akupunktur ist nicht wirksam bei der Raucherentwöhnung
ter Riet 90/89 (454;456)	A. Raucherentwöhnung	Sham, keine Behandlung, Standard, andere Akupunktur	A. 13 CT, 2 CCT	j-t-j-j-n	A. 3 von 15 Studien positiv	Keine sichere Evidenz für eine Wirksamkeit bei den drei Indikationen
	B. Alkoholentzug		B. 1 RCT, 1 CCT		B. 2 von 2 Studien positiv	
	C. Heroinentzug		C. 5 CCT		C. 3 von 5 Studien positiv	

Abkürzungen siehe Tabelle 17

Tabelle 20:
Systematische Übersichtsarbeiten zu klinischen Studien zur Akupunktur bei Übelkeit und Asthma

Autor Jahr	&	Indikation	Kontrolle	Studien	Method. 1-2-3- 4-5	Ergebnisse	Schlussfolgerungen
Übelkeit							
Lee (270)	99	Postop. Übelkeit	Sham, keine Behandlung, Standard	19 RCT	j-t-j-j-j	RR Früherbrechen im Vgl. zu Antiemetika 0,89 (95%KI 0,47- 1,67) und zu Sham 0,47 (0,34- 0,64), verzögertes Erbrechen 0,81 (0,46-1,42)	Akupunktur effektiv bei Erwachsenen, offensichtlich jedoch nicht bei Kindern
Aikins- Murphy 99** (7)		Schwanger- schafts- übelkeit	Sham, keine Behandlung	7 RCT	j-t-n-j-n	6 von 7 P6-Akupressur Studien positiv, 1 Studie zur P6-Elektro- stimulation positiv	Vorhandene Evidenz positiv aber nicht überzeugend
Jewell (214)	98*	Schwanger- schafts- übelkeit	Sham, keine Behandlung	4 RCT	j-j-j-j-j	3 von 4 P6-Akupressurstudien positiv, beste Studie negativ	Vorhandene Evidenz positiv aber nicht überzeugend
Vickers (474) (P6- Stimu- lation)	96 (474)	Übelkeit postop., in Schwanger., bei Chemoth.	Sham, keine Behandlung, Standard	33 CCT	j-j-j-j-n	27 von 33 Studien positiv, 11 von 12 RCT vs. Sham positiv	Außer nach Narkoseinduktion scheint P6-Stimulation antiemetisch effektiv
Asthma							
Linde (280)	98	Asthma (nicht akute Anfälle)	Sham	7 RCT	j-j-j-j-n	Studien sehr heterogen, 2 positiv, 5 ohne klaren Effekt	Evidenz widersprüchlich, Empfehlungen für die Praxis nicht möglich
Linde (290)	96	Asthma	Sham, keine Behandlung, Standard	15 RCT	j-j-j-j-n	Studien sehr heterogen, Ergebnisse widersprüchlich	Evidenz nicht ausreichend, um zuverlässige Schluss- folgerungen ziehen zu können
Kleijnen (246) & ter Riet (452)	91 (246) & ter 89 (452)	Asthma	Unklar	9 RCT, 4 CCT	j-t-j-n-n	Studienqualität gering, Ergebnisse widersprüchlich	Vorhandene Evidenz kann eine Wirksamkeit nicht belegen

Abkürzungen siehe Tabelle 17

Systematische Übersichtsarbeiten zu einer Reihe weiterer Indikationen bzw. Themen sind in Tabelle 21 zusammengestellt. Beim Tinnitus sind die Ergebnisse der vorliegenden Studien wenig vielversprechend (93;360), bei der Unterstützung der Gewichtsreduktion bzw. Appetitverminderung widersprüchlich (108). Der bis zum Abschluss unserer Literatursuche einzig vorliegende systematische Review zur Schlaganfallreduktion bewertete die vorhandene Evidenz als vielversprechend (126). Während der Erstellung dieses Textes wurde jedoch von der gleichen Arbeitsgruppe eine aktualisierte Version des Reviews publiziert (359), in den neuere, besser kontrollierte und verblindete Studien vs. Sham-Akupunktur einbezogen wurden, die die positiven Ergebnisse der offenen Studien nicht bestätigen konnten.

Schließlich liegt ein Review zur Akupressur bei verschiedenen Indikationen (170) und zur Akupunktur bei zahnmedizinischen Indikationen vor (389).

Tabelle 21:
Systematische Übersichtsarbeiten zu klinischen Studien zur Akupunktur bei verschiedenen Indikationen

Autor Jahr	& Indikation	Kontrolle	Studien	Method. 1-2-3- 4-5	Ergebnisse	Schlussfolgerungen
Park 2000 (360)	Tinnitus	Sham, Standard	6 RCT	j-j-j-j-n	2 offene Studien positiv, 4 Blindstudien ohne signifikante Unterschiede	Keine Evidenz aus sorgfältigen Studien für eine Wirksamkeit bei Tinnitus
Dobie 99 * (93)	Tinnitus	Sham, Standard	6 RCT	t-t-t-t-n	Keine signifikanten Effekte. Patienten in cross-over-Studien meist mit Präferenz für Akupunktur	Keine spezifischen Schlussfolgerungen zur Akupunktur (generell: keine Behandlung belegt wirksam)
Ernst 97 (108)	Gewichts- und Appetits- reduktion	Sham	4 RCT	?-j-j-j-j	2 Studien mit positivem Effekt v.a. auf Appetit, 2 bessere Studien ohne positives Ergebnis	Keine Evidenz aus sorgfältigen Studien für eine Wirksamkeit bzgl. Gewichtsreduktion
Ernst 96 (126)	Schlaganfall- rehabilitation	Nur Reha	5 RCT, 1 CCT	j-j-n-j-n	Alle Studien beschreiben positiven Effekt auf Funktion, jedoch zahlreiche methodische Schwächen	Evidenz ermutigend aber nicht überzeugend
Harris 97 (nur Aku- pressur) (170)	verschiedene	Sham, keine Behandlung, Standard	23 klinische Studien	t-n-j-n-n	P6-Akupressur bei Nausea belegt, ansonsten wenig gute Studien	Keine klaren Schlussfolgerungen gezogen
Rosted 98 (389)	Zahnmedizin (v.a. temp.- mandibuläre Dysfunktion)	Sham, keine Behandlung, Standard	15 RCT	j-t-j-n-n	Relevante Schwächen in den meisten Studien. 11 von 15 positiv	Evidenz meist nicht überzeu- gend, bei temp.-mand. Dysfunkt. und Gesichtsschmerz Wirksamkeit wahrscheinlich

Abkürzungen siehe Tabelle 17

4.3.2 Zusammenstellung der systematischen Übersichtsarbeiten zur Phytotherapie

Zur Phytotherapie wurden insgesamt 79 potentiell relevante Übersichtsarbeiten durch den Screeningprozess vorselektiert. Davon erfüllten 58 (publiziert in 65 Veröffentlichungen bzw. Berichten) die Einschlusskriterien. Dabei wurden aktualisierte Versionen eines Reviews von der gleichen Autorengruppe wiederum als ein Review gezählt. 11 Reviews wurden ausgeschlossen, da sie nicht als systematisch klassifiziert wurden (Nichterfüllung von Einschlusskriterium 2) (53;85;148;220;247;249;316;493;498;521), 5 bearbeiteten Studien zu einzelnen Inhaltsstoffen von Pflanzenextrakten (121;313;314;374;510) und weitere 4 aus unterschiedlichen Gründen (1 indikationsspezifischer Review mit weniger als 4 Studien zur Phytotherapie (98), 1 Review zur Korrelation von Knoblauchaufnahme in der Nahrung und Krebs (109), sowie 2 Reviews zu nicht rein phytotherapeutischen Präparaten (221;526)).

Mehr als die Hälfte der eingeschlossenen Übersichtsarbeiten waren zu Ginkgo-, Hypericum- oder Knoblauchpräparaten. Nicht weniger als 13 liegen zu Ginkgoextrakten (*Ginkgo biloba*) vor (siehe Tabelle 22 23). Davon behandeln 7 (mit insgesamt 15 Primärstudien; einer der Reviews

Tabelle 22:
Systematische Übersichtsarbeiten zu Ginkgo biloba bei Claudicatio intermittens

Autor & Jahr	Indikation	Intervention	Kontrolle	Studien	Method. 1-2-3- 4-5	Ergebnisse	Schlussfolgerungen
Pittler (376)	2000 Claudicatio intermittens	Ginkgo	Placebo	8 RCT	j-j-j-j-j	Zunahme der schmerzfreien Gehstrecke nach 12-24 Wochen 34 m (95%KI 26-43 m)	Kausaler Effekt von fraglicher klinischer Relevanz grundsätzlich belegt
Moher (330)	2000 Claudicatio intermittens	Ginkgo*	Placebo	5 RCT	j-j-j-n-j	Zunahme der schmerzfreien Gehstrecke nach 24 Wochen 32 m (95%KI 14-50 m)	Die inkonsistenten Ergebnisse der kleinen Studien lassen keine Schlussfolgerungen zu
Ernst (106)	96 Claudicatio intermittens	Ginkgo	Placebo, Standard	10 RCT/ CCT	t-t-n-n-n	Zunahme der schmerzfreien Gehstrecke vs. Placebo 24 - 160 m. Ähnlich effektiv wie Vergleichspräparate	Vorhandene Evidenz vielversprechend – weitere hochwertige Studien notwendig
Schneider (401)	92 Claudicatio intermittens	Ginkgo Extrakt EGb 761	Placebo, Standard	9 RCT/ CCT	?-n-n-j-j	Standardisierter Effekt d = 0,75 (95%KI 0,44-1,07) über Placebo	Effektivität im Vgl. zu Placebo eindeutig belegt
Letzel (277)	92 Claudicatio intermittens	Ginkgo Extrakt EGb 761	Ginkgo vs. Plac., Pentoxif. vs. Plac.	5 RCT Ginkgo, 9 RCT Pentoxif	?-t-n-j-j	Zunahme der Gehstrecke: 45% über Placebo für Ginkgo und 57% für Pentoxifyllin	Ginkgo Extract EGb761 effektiver als Placebo und ähnlich effektiv wie Pentoxifyllin
Kleijnen (240)	91 Claudicatio intermittens	Ginkgo	Ginkgo vs. Plac., Pentoxif. vs. Plac.	15 CCT Ginkgo, 5 CCT Pentoxif	j-j-j-n-n	Viele Studien mit mäßiger Qualität. Alle Studien positiv. Evidenz vergleichbar wie für Pentoxifyllin	Vorhandene Evidenz vielversprechend – weitere hochwertige Studien notwendig
Weiss (496)	91 Claudicatio intermittens	Ginkgo Extrakt EGb761	Placebo	8 CCT	?-t-t-n-n	Alle 4 interpretierbaren Studien positiv	Effektivität im Vgl. zu Placebo eindeutig belegt

Legende siehe Tabelle 17

behandelte ebenfalls Studien zur zerebrovaskulären Insuffizienz) das Thema Claudicatio intermittens (106;240;277;330;376;401;496). Eine Vergrößerung der schmerzfreien Gehstrecke unter Ginkgotherapie scheint grundsätzlich gut belegt, die klinische Relevanz dieser Effekte wurde jedoch zum Teil in Frage gestellt (330;376). Die 5 Reviews zu verschiedenen demenziellen Syndromen ((122;191;241;349;496), insgesamt ca. 50 Primärstudien, siehe Tabelle 23) ziehen grundsätzlich alle positive Schlussfolgerungen. Ein großer Teil der älteren Studien wurde jedoch mit Patienten mit leichteren kognitiven Einschränkungen durchgeführt. Zusätzliche Studien sind notwendig, um eine Beurteilung zu erlauben, ob Ginkgoextrakte klinisch relevante Effekte bei fortgeschritteneren bzw. schwerwiegenderen Demenzformen haben.

Tabelle 23:
Systematische Übersichtsarbeiten zu Ginkgo biloba bei demenziellen Syndromen und anderen Erkrankungen

Autor & Jahr	Indikation	Intervention	Kontrolle	Studien	Method. 1-2-3- 4-5	Ergebnisse	Schlussfolgerungen
Ernst (122)	99 Demenz	Ginkgo	Placebo	9 RCT	j-j-j-j-n	Vorliegende Studien überwiegend mit positiven Effekten	Vielversprechende Ergebnisse – größere Studien notwendig
Oken (349)	98 M. Alzheimer	Ginkgo	Placebo	4 RCT	j-j-n-j-j	Signifikanter Effekt über Placebo bzgl. Kognition (Hedges g = 0,41, 95%KI 0,22-0,61)	Die klinische Relevanz der berichteten Effekte sollte in weiteren Studien überprüft werden
Hopfenmüller 94 (191)	Zerebrovask. Insuffizienz	Ginkgo Extrakt 1370	Placebo LI	10 RCT, 1 CCT	n-n-n-j-j	OR Globalbewertung 1,98 (95%KI 1,39-2,57) zugunsten von Ginkgo	Ginkgoextrakt einer Placebobehandlung überlegen
Kleijnen (241)	92 Zerebrovask. Insuffizienz	Ginkgo	Ginkgo vs. Plac. Hydergin vs. Plac.	40 CCT Ginkgo, 4 CCT Hyderg.	j-j-j-n-n	Studienqualität meist schlecht, alle Ergebnisse positiv. Evidenz ähnlich wie für Hydergin	Ginkgoextrakte scheinen effektiv zu sein – weitere hochwertige Studien notwendig
Weiss (496)	91 Zerebrovask. Insuffizienz	Ginkgo Extrakt EGb761	Placebo	17CCT	?-t-t-n-n	10 von 12 interpretierbaren Studien positiv	Effektivität im Vgl. zu Placebo eindeutig belegt
Ernst (125)	99 Tinnitus	Ginkgo	Placebo, andere	5 RCT	j-j-j-j-n	In 3 Studien Ginkgo besser als Placebo, in 1 kein Unterschied	Ergebnisse vielversprechend
Evans (129)	2000 Macula-degeneration	Ginkgo	Placebo	1 RCT	j-j-j-j-n.a.	Kleine Studie mit positivem Ergebnis	Vorliegende Evidenz nicht ausreichend für zuverlässige Schlussfolgerungen

Legende siehe Tabelle 17

Auch zur Behandlung des Tinnitus mit Ginkgoextrakten liegt eine systematische Übersichtsarbeit vor. Die Ergebnisse wurden als vielversprechend bewertet (125). Inzwischen wurde allerdings eine größere randomisierte Studie veröffentlicht, die die positiven Ergebnisse nicht bestätigen konnte (96). Ein weiterer Review zu Ginkgo zur Maculadegeneration umfasste nur eine einzige, kleinere Studie (129), aus der sich keine eindeutigen Schlussfolgerung ableiten lassen.

Die Wirksamkeit von Hypericumkrautextrakten (*Hypericum perforatum*, syn. Johanniskraut) bei Patienten mit depressiven Verstimmungen im Vergleich zu Placebo und Standardantidepressiva wurde in 9 systematischen Reviews untersucht (siehe Tabelle 24, 11 Publikationen (104;139;142;235;285;286;337;441;483;485;509), insgesamt 29 Primärstudien). Primär aufgrund unterschiedlicher Einschlusskriterien differieren die jeweiligen Studiensammlungen in erheblichem Maße. Die Schlussfolgerungen waren jedoch in allen Fällen sehr ähnlich: Bei leichten bis mittelschweren Depressionen ist die Wirksamkeit von Hypericumextrakten im Vergleich zu Placebo nachgewiesen. Die Äquivalenz mit (angemessen dosierten) Standardantidepressiva ist noch nicht sicher zu beurteilen. Hier werden weitere Studien benötigt. Im Vergleich zu älteren Antidepressiva sind die Nebenwirkungsraten unter Hypericumbehandlung deutlich niedriger.

8 systematische Übersichtsarbeiten (in 9 Publikationen) liegen zum Einfluss von Knoblauchpräparaten (*Allium sativum*) auf kardiovaskuläre Risikofaktoren ((239;244;268;342;411;412;443;491) insgesamt ca. 50 Primärstudien) und zur Atherosklerose in den Beinen vor ((211), siehe Tabelle 25). Ein mäßig ausgeprägter, kurzzeitiger cholesterinsenkender Effekt scheint für eine Reihe von Knoblauchpräparaten grundsätzlich belegt. Langzeiteffekte sind nicht sicher belegt. Langzeitauswirkungen auf kardiovaskuläre Ereignisse und Mortalität wurden bisher nicht untersucht. Blutdrucksenkende Effekte sind, wenn überhaupt, eher schwach ausgeprägt. Die Ergebnisse bzgl. der fibrinolytischen Aktivität und der Thrombozytenaggregationshemmung erscheinen grundsätzlich vielversprechend, aber nicht ausreichend abgesichert (268). Ein besonderes Problem ist die mangelnde Vergleichbarkeit unterschiedlicher Knoblauchzubereitungen.

Tabelle 24:
Systematische Übersichtsarbeiten zu Johanniskrautextrakten (*Hypericum perforatum*)

Autor Jahr	&	Indikation	Intervention	Kontrolle	Studien	Method. 1-2-3- 4-5	Ergebnisse	Schlussfolgerungen
Gaster 2000 (142)		Depression	Hypericum	Placebo, Standard	8 RCT	t-j-t-j-n	4 placebokontr. Studien mit positiven Effekten, in 4 Vergleichen mit Antidepressiva tendenziell unterlegen	Hypericum wohl besser als Placebo, Äquivalenz mit Antidepressiva unklar
Williams 2000 & Mulrow 98* (337)		Depression	Hypericum	Placebo, Standard	14 RCT	j-j-n-j-j	Responderratio 1,9 (95%KI 1,2-2,8) vs. Placebo und 1,2 (1,0-1,4) vs. Antidepressiva	Hypericum effektiver als Placebo, Äquivalenz mit Antidepressiva nicht sicher beurteilbar
Kim (235)	99	Depression	Hypericum	Placebo, Standard	6 RCT	t-j-j-j-j	Responderratio 1,48 (95%KI 1,03-1,92) vs. Placebo und 0,98 (0,67-1,28) vs. Antidepressiva	Hypericum effektiver als Placebo, ähnlich wie niedrig dosierte Antidepressiva; Studienqualität mäßig
Stevinson 99 (441)		Depression	Hypericum	Placebo, Standard	6 RCT	j-j-j-j-n	Nur Studien publiziert nach Linde 96; Überlegenheit vs. Placebo, ähnliche Effekte wie Antidepressiva	Neue Studien bestätigen ältere; Äquivalenz mit Antidepressiva immer noch unsicher
Linde 98 & 96 (285;286)		Depression	Hypericum	Placebo, Standard	27 RCT	j-j-j-j-j	Responderratio 2,47 (95%KI 1,69-3,61) vs. Placebo und 1,01 (0,87-1,16) vs. Antidepressiva	Hypericum effektiver als Placebo, Äquivalenz mit Antidepressiva nicht sicher beurteilbar
Volz (483)	97	Depression	Hypericum	Placebo, Standard	15 CCT	t-t-n-n-n	Studien vs. Placebo meist positiv; ähnliche Effekte wie niedrig dosierte Standard-Antidepressiva	Bei leichter bis mäßiger Depression Therapieversuch mit Hypericum angemessen
Ernst (104)	95	Depression	Hypericum	Placebo, Standard	11 RCT	j-j-j-j-n	Mehrheit der 8 placebokontrollierten Studien positiv. 3 Studien vs. Standard ohne Unterschiede	Hypericum Placebo überlegen und ähnlich effektiv wie Standard-Antidepressiva
Volz 2000 (485)		Leichte bis mittelschw. Depression	Hypericum	Fluoxetin	17 + 9 CCT	n-j-n-j-n	Vgl. von Responderaten in Hypericum- (53%) und Fluoxetinstudien (55%); kein direkter Vergleich	Ähnliche Responderaten – problematisch zu interpretieren
Friede (139)	98	Angst bei depressiven Verstärkungen	Hypericum	Placebo, Amitriptylin	8 RCT	?-j-j-j-n	Rückgang der Angstsymptomatik unter Hypericum ausgeprägter im Vgl. zu Placebo	Hypericum effektiv zur Behandlung der Angstsymptomatik bei depressiven Patienten

Legende siehe Tabelle 17

Ein ähnliches Problem besteht auch bei Echinaceaextrakten, die häufig in Kombination mit anderen Extrakten verwendet werden. Hier liegen 3 systematische Übersichtsarbeiten (zu insgesamt 30 Primärstudien) vor ((29;320;322), siehe Tabelle 26). Vielversprechend erscheint hier die Evidenz bzgl. der Frühbehandlung von Erkältungskrankheiten, d.h. beim Auftreten der ersten Infektsymptome. Bei der eigentlichen Infektophylaxe (Einnahme von Echinaceapräparaten über längere Zeiträume) erscheinen die Effekte allenfalls gering, bei der Behandlung späterer Infektstadien sind die Ergebnisse ebenfalls nicht überzeugend.

Tabelle 25:
Systematische Übersichtsarbeiten zu Knoblauchpräparaten (*Allium sativum*)

Autor Jahr	& Indikation	Intervention	Kontrolle	Studien	Method. 1-2-3- 4-5	Ergebnisse	Schlussfolgerungen
Lawrence 2000 (268)	Kardiovask. Risiko- faktoren	Knoblauch	Meist Placebo	45 RCT	j-j-j-j-j	37 Studien mit kleinen, aber konsistenten Kurzzeiteffekten bzgl. Cholesterin. Ergebnisse zu RR-Senkung widersprüchlich, vielversprechend bzgl. Gerinnungshemmung	Nur Untersuchungen zu Surrogatvariablen. Moderate Kurzzeiteffekte auf Cholesterin können als belegt gelten. Äußerst unterschiedliche Präparate
Stevinson 2000 (443)	Hyperchol- esterinämie	Knoblauch	Placebo	13 RCT	j-j-j-j-j	Cholesterinreduktion über Placebo 0,41 (95%KI -0,66 to -0,15) mmol/l; in besten Studien 0,11 (-0,30 to 0,08)	Kleiner signifikanter Effekt von fraglicher klinischer Relevanz
Silagy 94 & Neil 96 (342;411)	Cholesterin- senkung	Knoblauch	Placebo	16 RCT	j-t-j-j-j	Cholesterinreduktion über Placebo 0,65 (95%KI 0,53-0,76) mmol/l	Meta-Analyse suggeriert signifikanten Effekt, Reviewer sind jedoch skeptisch
Warshafsky 93 (491)	Cholesterin- senkung	Knoblauch	Placebo	5 RCT	t-j-j-j-j	Cholesterinreduktion über Placebo 0,59 (95%KI 0,44-0,74) mmol/l	Vorhandene Evidenz spricht dafür, dass Knoblauch eine sinnvolle Therapieoption sein könnte
Silagy 94 (412)	Blutdruck- senkung	Knoblauch- pulver (Kwai)	Placebo, andere Behandl.	8 RCT	j-t-j-j-j	RR-Reduktion über Placebo: syst. 7,7 (95%KI 4,3-11,0), diast. 5,0 (2,9-7,1) mm Hg	Knoblauch möglicherweise blutdrucksenkend wirksam – weitere Studien notwendig
Kleijnen 91 (239)	Kardiovask. Risiko- faktoren	Knoblauch	Placebo	18 CCT	t-t-j-j-n	Methodische Qualität meist mäßig. Meist positive Ergebnisse, Effektgrößen jedoch heterogen	Keine klare Schlussfolgerung gezogen
Kleijnen 89 (244)	Kardiovask. Risiko- faktoren	Knoblauch und Zwiebeln	Unklar	10 RCT, 8 CCT	j-t-n-j-n	Alle Studien mit großen Qualitätsmängeln. Positive Effekte bei frischem Knob., widersprüchlich bei Präp.	Evidenz nicht ausreichend für zuverlässige Schlussfolgerungen
Jepson 97 (211)	Bein-Athero- sklerose	Knoblauch	Placebo	1 RCT	j-j-j-j- n.a.	Keine signifikanten Unterschiede zwischen Gruppen bzgl. Gehstrecke	Evidenz nicht ausreichend um Schlussfolgerungen zu ziehen

Legende siehe Tabelle 17

Zwei systematische Reviews liegen zur Gabe von Preisselbeersaft zur Prävention bzw. Therapie von Harnwegsinfekten vor ((212;213), siehe Tabelle 26). Während bzgl. des prophylaktischen Einsatzes vielversprechende Studienergebnisse publiziert wurden, gibt es zur Behandlung keine Evidenz aus kontrollierten Studien: der entsprechende Review ist „leer“, d.h. keine einzige Studie entsprach den Einschlusskriterien.

Die v.a. in deutschsprachigen Ländern weit verbreitete Misteltherapie ist seit 1994 nicht mehr in systematischen Reviews untersucht worden. Die bis dahin vorliegenden Übersichten ((233;234;243), siehe Tabelle 26) fanden zwar überwiegend positive Primärstudien, die Studienqualität war jedoch unbefriedigend. Außerdem konnte in der methodisch besten Studie kein positiver Effekt auf die Überlebenszeit festgestellt werden. Für die Behandlung des Colon irritabile mit Pfefferminzöl sind die in zwei Reviews (207;373) zusammengefassten vorliegenden Studien in der Mehrheit ebenfalls positiv, aufgrund methodischer Schwächen und kurzer Beobachtungszeiträume kann die Evidenz jedoch noch nicht als überzeugend angesehen werden.

Für die Wirksamkeit von Sägepalmenextrakten (*Serenoa repens*) bei der benignen Prostatahyperplasie gibt es gute Belege (49;511;512). Einzelne systematische Übersichtsarbeiten liegen zu einer ganzen Reihe weiterer Zubereitungen aus Pflanzen vor (siehe Tabellen 27 und 28):

- *Aloe vera* bei verschiedenen Erkrankungen (480);
- Artischockenextrakt (*Cynara scolymus*) zur Cholesterinsenkung (371);
- Nachtkerzenöl (*Oenothera biennis*) bei Neurodermitis (336);
- Mutterkraut (*Tanacetum parthenium*, feverfew) bei Migräne (481);
- Ingwerwurzel (*Zingiber officinalis*) bei Übelkeit und Erbrechen (123);
- Ginsengwurzel (*Panax ginseng*) bei verschiedenen Indikationen (482);
- Rosskastaniensamen (*Aesculus hippocastanum*) bei chronisch venöser Insuffizienz (372);
- Kavawurzel (*Piper methysticum*) bei Angststörungen (375);
- Mariendistel (*Silybum marianum*) bei Lebererkrankungen (269);
- Einem Kombinationspräparat aus *Populus tremula*, *Fraxinus excelsior* und *Solidago virgaurea* bei Schmerzen in Zusammenhang mit Störungen des Bewegungsapparates (115);
- *Secale cereale* Pollenextrakt bei benigner Prostatahyperplasie (302;513);
- Teebaumöl (*Melaleuca alternifolia*) bei verschiedenen Hauterkrankungen (118);
- Baldrianwurzel (*Valeriana officinalis*) bei Schlafstörungen (442);

Zur Behandlung von Beschwerden von stillenden Frauen (Milchstau, Mastalgie, Brustentzündung) mit Kohlblättern liegt ein Review vor (386). Dies ist der einzige Review, in dem eine nicht vermarkt看bare phytotherapeutische Anwendung untersucht wurde. Jedoch konnte nur eine einzige Studie identifiziert werden. Obwohl die Ergebnisse vielversprechend sind, erscheinen angesichts

des geringen finanziellen Interesses weitere Studien zu Kohlblätteranwendungen eher unwahrscheinlich.

Tabelle 26:
Systematische Übersichtsarbeiten zu Extrakten verschiedener Pflanzen (zu denen mind. je 2 Reviews vorliegen)

Autor Jahr	& Indikation	Intervention	Kontrolle	Studien	Method. 1-2-3- 4-5	Ergebnisse	Schlussfolgerungen
Echinacea (Echinacea purpurea, angustifolia und pallida)							
Barrett 99 (29)	Infekte der oberen Luftwege	Echinacea (incl. Kombinationspräparate)	Placebo	13 RCT	j-t-j-j-n	Studienqualität mäßig. Alle 4 Präventionsstudien nur mit Trend, 8 von 9 Behandlungsstudien positiv	Präventiver Effekt längerer Einnahme nicht belegt, Ergebnisse bzgl. kurativem Effekt vielversprechend
Melchart 99 (320)	Erkältung	Echinacea (incl. Kombinationspräparate)	Placebo, keine Behandl.	16 RCT	j-j-j-j-t	Positive Effekte bei Frühbehandlung, gering bei Prävention und Behandlung ausgeprägter Infekte	Einzelne Echinaceaextrakte zur Frühbehandlung wohl wirksam; Evidenz nicht ausreichend für Empfehlung
Melchart 94 (322)	Immunstimulation	Echinacea (incl. Kombinationspräparate)	Placebo, keine Behandl.	18 RCT, 8 CCT	j-j-j-j-n	Studien meist mit mäßiger Qualität und positiven Ergebnissen	Einzelne Echinaceaextrakte wohl wirksam; Evidenz nicht ausreichend für Empfehlung
Preisselbeersaft							
Jepson 98 (213)	Prävention Harnwegsinfekte	Preisselbeersaft	Placebo	4 RCT	j-j-j-j-n	In 3 von 4 Studien Wirksamkeit bzgl. mind. eines relevanten Parameters	Evidenz nicht ausreichend – weitere Studien notwendig
Jepson 98 (212)	Behandlung Harnwegsinfekte	Preisselbeeren		0 RCT	j-j-n.a.-n.a.-n.a.	Keine Studie gefunden, die den Einschlusskriterien entsprach	-
Mistel (Viscum album)							
Kleijnen 94 (243)	Krebs	Mistel	Placebo, keine Zusatzbehandl.	11 CCT	j-j-j-n-n	Studienqualität meist mäßig. Alle Studien bis auf die beste zeigen Überlebensvorteil unter Misteltherapie	Evidenz reicht nicht aus, um Misteltherapie außerhalb klinischer Studien zu empfehlen
Kiene 89 (233;234)	Krebs	Mistel	Keine Zusatzbeh., keine Kontrolle	2 RCT, 33 CCT, 11 andere	j-n-n-j-n	Studienqualität meist mäßig. 9 von 12 interpretierbaren Studien zeigen Überlebensvorteil unter Misteltherapie	Vorhandene Evidenz positiv
Pfefferminze (Mentha piperita)							
Jailwala 2000* (207)	Colon irritabile	1. Pfefferminzöl 2. chin. Kräuter	Placebo	1. RCT 2. RCT	3 t-j-j-n-n 1	1 von 3 Pfefferminzstudien und Studie zu chin. Kräutern mit positiven Ergebnissen	Evidenz in beiden Fällen nicht ausreichend
Pittler 98 (373)	Colon irritabile	Pfefferminzöl	Placebo, andere Behandl.	8 RCT	j-j-j-j-j	Studienqualität mäßig. Globalbeurteilung unter Pfefferminzöl besser als unter Placebo	Evidenz vielversprechend aber nicht ausreichend

Legende siehe Tabelle 17

Tabelle 27:
Systematische Übersichtsarbeiten zu Zubereitungen verschiedener Pflanzen

Autor Jahr	&	Indikation	Intervention	Kontrolle	Studien	Method. 1-2-3- 4-5	Ergebnisse	Schlussfolgerungen
Boyle 2000 (49)		Benigne Prostata- hyperplasie	Permixon® (Sägepal- menextrakt)	Placebo, Standard	11 RCTs, 2 UKS	?-n-n-j-j	Anstieg Flowwerte 2,20 (95%KI 1,20-3,20) ml/s über Placebo; Nykturie signifikant vermindert	Trotz einiger Einschrän- kungen deutliche Evidenz für Wirksamkeit vorhanden
Wilt 2000 & 98 (511;512)		Benigne Prostata- hyperplasie	Sägepalme	Placebo, Standard	19 RCT	j-j-j-j-j	Überlegenheit vs. Placebo bzgl. Nykturie, Urinflow, Patientenurteil; ähnlich ef- fektiv wie Finasterid	Vorhandene Evidenz spricht für Wirksamkeit – weitere Studien notwendig
Vogler 99 (480)		Verschie- dene	Aloe vera	Placebo, Standard, keine Beh.	6 RCT, 4 CCT	j-j-j-j-n	Positive Studien bei Genital- herpes, Psoriasis, Hyper- lipidaemie, Diabetes; Wundheilung kontrovers	Vielpersprechende, aber nicht ausreichende Evidenz
Pittler 98 (371)		Cholesterin- senkung	Artischok- kenblatt- extrakt	Placebo	1 RCT	j-j-j-n-n	Positiver Effekt über Placebo nur bei Patienten mit Werten > 210 mg/dl	Weitere Studien notwendig
Morse 89 (336)		Neurodermi- tis	Nachtker- zenöl	Placebo	9 CCT	?-n-n-j-j	Signifikant besser als Placebo für verschiedene Parameter	Keine Schlussfolgerung gezogen
Vogler 98 (481)		Migräne	Mutterkraut	Placebo	5 RCT	j-j-j-j-n	Mehrheit der Studien mit Wirksamkeit über Placebo	Wirksamkeit nicht ausreichend sicher belegt
Ernst 2000 (123)		Übelkeit und Erbrechen	Ingwer- wurzel	Placebo, Metoclo- pramid	6 RCT	j-j-j-j-t	2 von 3 Studien postop. positiv, Studien zu See- krankheit, Schwangersch. und Chemotherapie pos.	Evidenz vielversprechend aber nicht ausreichend
Vogler 99 (482)		Verschie- dene	Ginseng- wurzel	Placebo, Standard (1 Studie)	16 RCT	j-t-j-j-n	Ergebnisse widersprüchlich: körperl. Leistung (7 Stud.), psychol. Funktion (5), Im- munmodulation (2); positiv: Diabetes (1), Herpes (1)	Wirksamkeit von Ginsengwurzel bei keiner der geprüften Indikationen sicher belegt
Pittler 98 (372)		Venöse Insuffizienz	Rosskasta- nie	Placebo, Standard	13 RCT	j-j-j-j-n	Signifikant wirksamer als Placebo und ähnlich wie andere Therapien	Vorhandene Evidenz positiv aber weitere Studien notwendig
Pittler 2000 (375)		Angst- störungen	Kava	Placebo	7 RCT	j-j-j-t-t	Alle Studien positiv. Meta- Analyse von 3 Studien ebenfalls signifikant positiv	Vorhandene Evidenz positiv. Weitere Studien notwendig
Lawrence 2000 (269)		Leberer- krankungen	Mariendistel	Placebo, Standard, keine Beh.	33 RCT, 1 CCT	j-j-j-j-j	Verschiedene Lebererkr. untersucht; Qualität meist mäßig, Ergebnisse inkonsistent	Wirksamkeit nicht belegt. Positive Effekte am ehesten auf Leberfunktion (Trans- aminasen)

Legende siehe Tabelle 17

Tabelle 28:
Systematische Übersichtsarbeiten zu Zubereitungen verschiedener Pflanzen

Autor Jahr	&	Indikation	Intervention	Kontrolle	Studien	Method. 1-2-3- 4-5	Ergebnisse	Schlussfolgerungen
Ernst (115)	99	Schmerzen im Bew.app.	Phytodolor® (Populus, Fraxinus, Solidago)	Placebo, Standard	10 RCT	j-t-j-j-n	Überlegenheit im Vgl. zu Placebo und ähnliche Effekte wie nichtsteroidale Antiphlogistika	Vorhandene Evidenz positiv
MacDonald / Wilt (302;513)	2000	Benigne Prostata-hyperplasie	Secale cereale Pollen-extrakt	Placebo, Standard	4 RCT	j-j-j-j-j	Signifikant besser als Plac. für subjektive (nicht objektive) Symptome; ähnlich eff. wie Tadenan und Paraprost	Vorhandene Evidenz vielversprechend – weitere Studien notwendig
Ernst (118)	2000	Hauterkrankungen	Teebaumöl	Placebo, Standard	4 RCT	j-j-j-j-n	2 Studien vs. Placebo positiv, 3 vs. andere Behandlung ohne Unterschiede	Evidenz vielversprechend aber bisher nicht ausreichend
Stevinson (442)	2000	Schlafstörungen	Baldrianwurzel	Placebo	9 RCT	j-j-j-j-n	Sehr heterogene Studien mit zum Teil widersprüchlichen Ergebnissen	Vorhandene Evidenz nicht eindeutig, aber grundsätzlich vielversprechend – weitere Studien notwendig
Renfrew (386)	84	Milchstau, Mastalgie	Kohlblätter	Übliche Behandlung	1 RCT	j-j-j-j-n	Unter den Frauen, die Kohlblätterumschläge machten, mußten weniger abstillen	Weitere Forschung wünschenswert
Armstrong (20)	99	Neurodermitis	Chinesische Kräuter	Placebo	2 RCT	j-j-j-j-n	2 positive Studien der gleichen Forschungsgruppe, keine intent-to-treat Analyse	Vorhandene Evidenz vielversprechend – weitere Studien notwendig
Ernst (114)	97	Blutzuckersenkung	Keine Einschränkung	Placebo, keine Beh., keine Kontrolle	7 RCT, 4 CCT, 10 UCS	j-t-n-j-n	Qualität meist gering. Blutzuckersenkende Wirkungen bei einer Reihe von Extrakten beschrieben	Bisher keine überzeugenden Belege vorhanden. Weitere Forschung gerechtfertigt
Ernst (117)	2000	Analgetische und antiinfl. Behandlung	Keine Einschränkung	Placebo	18 RCT	j-j-j-j-n	Meist pos. Ergebnisse zu Nachtkerzen-, schwarzen Johannisbeer- und Borretschöl, Teufelskrallen, Weidenrinde, Mutterkraut und Kombipräparaten	Ergebnisse zum Teil sehr vielversprechend – weitere Forschung sehr wünschenswert

Legende siehe Tabelle 17

Eine feste Mischung von Kräutern der traditionellen chinesischen Medizin zur Behandlung der Neurodermitis war Thema eines systematischen Reviews, in den 2 Primärstudien (vom selben Forschungsteam) eingeschlossen wurden (20).

Schließlich liegen noch zwei weitere systematische Übersichtsarbeiten vor, die Studien zusammenfassen, in denen verschiedene Pflanzen auf ihre blutdrucksenkende (114) oder entzündungshemmende Wirksamkeit (117) untersucht worden waren. Besonders im zweiten Fall wurde eine Reihe vielversprechender Ergebnisse gefunden; weitere klinische Studien wurden hier als besonders sinnvoll erachtet.

4.3.3 Zusammenstellung der systematischen Übersichtsarbeiten zur Homöopathie

Von insgesamt 22 potentiell relevanten Übersichtsarbeiten zur Homöopathie erfüllten 18 (in 19 Publikationen) die Einschlusskriterien (siehe Tabellen 29 und 30). Vier Arbeiten wurden ausgeschlossen, da es sich entweder um Subgruppen- oder Reanalysen eingeschlossener Reviews handelte (110;124;219;287).

3 quantitative Meta-Analysen beschäftigten sich mit der grundsätzlichen Frage, ob für homöopathische Arzneimittel überhaupt ein Effekt über Placebo hinaus nachweisbar ist (76;278;489). Insgesamt 97 Primärstudien waren in mindestens einen der 3 Reviews eingeschlossen worden, in den größten 89. Die einzelnen Studiensammlungen differierten v.a. aufgrund unterschiedlicher Einschlusskriterien erheblich. Während zwei Meta-Analysen (76;278) eine signifikante Überlegenheit der homöopathischen Arzneimittel gegenüber Placebo fanden, war der Unterschied in der dritten (489) in der Hauptanalyse (mit einem random effects Modell) nicht signifikant. In einer weiteren (fixed effects) Analyse war der Unterschied dagegen auch in diesem Review hochsignifikant. Bereits mehrere Jahre vor der Publikation dieser Untersuchungen war eine quantitative Meta-Analyse versucht worden (299). Diese Arbeit beschäftigte sich aber primär mit den dabei auftretenden methodischen Problemen, erbrachte keine klaren Ergebnisse und zog keine inhaltlichen Schlussfolgerungen.

Zwei ältere systematische Übersichtsarbeiten bezogen sowohl placebokontrollierte Studien als auch Vergleiche mit anderen Medikamenten mit ein (185;245). Die Ergebnisse wurde in einem Vote Count als positiv oder negativ kategorisiert. Ein Review zog vorsichtig positive Schlussfolgerungen (245), der andere eher negative (185).

Zwei weitere krankheitsübergreifende Reviews schlossen ausschließlich Primärstudien ein, in denen eine klassisch homöopathische Intervention (d.h. individuell ausgewählte Arzneimittel) untersucht worden war. Einer dieser Reviews war darüberhinaus auf Vergleiche mit konventionellen medikamentösen Interventionen beschränkt (111). Lediglich 6 methodisch fehlerbehaftete Studien entsprachen diesen Einschlusskriterien, sodass eine zuverlässige Schlussfolgerung nicht möglich schien. Der zweite Review schloss auch eine größere Zahl placebokontrollierter Studien ein (283). Bei der meta-analytischen Auswertung ergab sich eine signifikante Überlegenheit gegenüber Placebo; wurde die Analyse auf die methodisch besten Studien beschränkt, war der Effekt jedoch deutlich kleiner und nicht mehr signifikant.

Tabelle 29:
Krankheitsübergreifende systematische Übersichtsarbeiten zur Homöopathie

Autor Jahr	& Indikation	Homöopath. Intervention/ Kontrolle	Studien	Method. 1-2-3-4-5	Ergebnisse	Schlussfolgerungen
Homöopathie allgemein krankheitsübergreifend						
Cucherat 2000 (76)	Keine Einschränk.	Keine Einschränk./ Placebo	17 RCT	j-j-j-n-j	Kombinierter p-Wert für einen Effekt über Placebo = 0,000036, für die besten Studien = 0,08	Es gibt Evidenz für einen Effekt über Placebo hinaus. Bessere Studien weniger positiv
Linde (278)	97 Keine Einschränk.	Keine Einschränk./ Placebo	89 RCT	j-j-j-j-j	OR alle Studien für Effekt über Placebo 2,45 (95%KI 2,05- 2,93), in besseren Studien 1,66 (1,33-2,08)	Evidenz für Effekt über Placebo vorhanden; keine überzeugende Evidenz bei einzelnen Erkrankungen
Walach (489)	97 Keine Einschränk.	Keine Einschränk./ Placebo oder Standard	41 RCT	j-t-j-j-j	Effektgröße g random effects = 0,259 (95%KI -0,319-0,837), fixed effects 0,295 (0,223- 0,366)	Effekt über Placebo nicht statistisch signifikant absicherbar
Lutz (299)	93 Keine Einschränk.	Keine Einschränk./ Placebo oder Standard	21 RCT/ CCT	?-n-j-j-t	Studienergebnisse widersprüchlich	Keine Schlussfolgerung (Dissertation hauptsächlich zu technischen Problemen der Meta-Analyse)
Kleijnen (245)	91 Keine Einschränk.	Keine Einschränk./ Placebo oder Standard	107 CCT	j-t-j-j-n	81 Studien mit positiven Ergebnissen; Qualität meist schlecht, jedoch viele Ausnahmen	Vorhandene Evidenz positiv aber nicht ausreichend für definitive Schlussfolgerungen
Hill (185)	90 Keine Einschränk.	Keine Einschränk./ Placebo oder Standard	40 RCT	n-t-j-j-n	Hälfte der Studien laut Autoren positiv, weitere 7 tendenziell positiv	Keine ausreichende Evidenz für eine Wirksamkeit
Individualisierte (klassische) Homöopathie krankheitsübergreifend						
Ernst (111)	98 Keine Einschränk.	Individuell/ Standard	3 RCT, 3 CCT	j-t-n-j-n	Alle Studien mit relevanten methodischen Schwächen; Ergebnisse unterschiedlich	Wirksamkeit klassischer Homöopathie im Vergleich zu konventionellen Medikamen- ten bisher nicht beurteilbar
Linde (283)	98 Keine Einschränk.	Individuell/ Placebo oder Standard	32 RCT	j-j-j-j-j	Responder RR vs. Placebo 1,62 (95%KI 1,17-2,23), für bessere Studien 1,12 (0,87- 1,44)	Vorhandene Evidenz grund- sätzlich positiv aber nicht überzeugend (Methodik, fehlende Replikationen)

Legende siehe Tabelle 17

Primär krankheitsbezogene Übersichtsarbeiten zur Behandlung mit homöopathischen Arzneimitteln liegen zum postoperativen Ileus (28), zum Muskelkater (116), bei rheumatischen Erkrankungen (200) und bei Asthma (281) vor (siehe Tabelle 30). Beim postoperativen Ileus hat die Mehrzahl der Studien einen signifikanten Effekt der homöopathischen Arzneimittel gezeigt, die methodisch beste und größte Studie jedoch nicht. In den anderen Fällen ist die vorliegende Evidenz wenig aussagekräftig.

Arnica wird in der Homöopathie sehr häufig bei der Behandlung von kleineren Verletzungen, Wunden etc. angewendet und ist das in klinischen Studien am häufigsten untersuchte Mittel. Zwei systematische Übersichtsarbeiten mit z.T. unterschiedlichen Einschlusskriterien (v.a. bzgl. des Einbezugs von Arnica in Kombination mit anderen Homöopathika bzw. in fixen homöopathischen Kombinationspräparaten) liegen vor ((119;301), insgesamt 37 Primärstudien). Die Ergebnisse der vorliegenden Studien sind widersprüchlich; der umfangreichere der beiden Reviews zieht jedoch etwas positivere Schlussfolgerungen.

Übersichtsarbeiten bzw. Meta-Analysen liegen darüberhinaus zur Gabe von isopathischen Nosoden bei allergischen Erkrankungen (448), für ein in Frankreich zur Infektbehandlung weit verbreitetes Nosodenpräparat (Oscillococcinum (476)), für die klassisch homöopathische Behandlung chronischer Kopfschmerzen (112) und für *Galphimia glauca* bei Heuschnupfen vor (300;508). Außer für die Behandlung von Kopfschmerzen sprechen in allen Fällen die vorliegenden Ergebnisse für eine Wirksamkeit. Im Falle der isopathischen Nosoden bei Allergie und *Galphimia glauca* bei Heuschnupfen sind jedoch unabhängige Replikationen der Studien unabdingbar, da die Primärstudien jeweils von einer Forschungsgruppe stammen.

Tabelle 30:
Systematische Übersichtsarbeiten zur Homöopathie mit beschränktem Krankheitsspektrum

Autor Jahr	& Indikation	Homöopath. Intervention/ Kontrolle	Studien	Method. 1-2-3-4-5	Ergebnisse	Schlussfolgerungen
Verschiedene homöopathische Arzneimittel bei definierten Indikationen						
Barnes (28)	97 Post-operativer Ileus	Verschiedene/ Placebo	4 RCT, 2 CCT	j-j-j-j-j	Zeit bis zur Wiederaufnahme der Darmtätigkeit unter hom. Beh. signifikant kürzer – beste Studie jedoch negativ	Vorhandene Evidenz positiv, verschiedene Probleme machen aber endgültige Schlussfolg. unmöglich
Ernst (116)	98 Muskelkater	Verschiedene/ Placebo	8 CCT	j-j-j-j-n	Studien meist mit großen Schwächen; bessere Studien ohne Effekt über Placebo hinaus	Vorhandene Evidenz spricht nicht für eine Wirksamkeit
Jacobs (200)	91* Rheumat. Erkrank.	Verschiedene Placebo	4 CCT	t-j-n-j-n	3 von 4 Studien positiv. Qualität der Studien schlecht	Keine spez. Schlussfolgerung zur Homöopathie (allg.: keine komplementäre Therapie gut belegt)
Linde (281)	98 Asthma	Verschiedene Placebo	3 RCT	j-j-j-j-n	Völlig unterschiedliche Interventionen. 2 Studien positiv	Vorhandene Evidenz lässt eine Beurteilung der Wirksamkeit bei Asthma nicht zu
Reviews zu Arnica bei verschiedenen Indikationen (meist Trauma)						
Lüdtke (301)	99 Keine Einschränkung.	Arnica/ Placebo, keine Behandlung	23 RCT, 14 CCT	j-j-j-n-n	Qualität meist gering. 13 von 35 Studien vs. Placebo mit signifikant positiven Ergebnissen, 10 mit Trend	Evidenz suggeriert, dass Arnica wirksam sein kann – weitere Studien notwendig
Ernst (119)	98 Keine Einschränkung.	Arnica/ Placebo	4 RCT, 4 CCT	j-j-j-j-n	2 Studien positiv, 2 mit positivem Trend. Studien meist mit relevanten Schwächen	Wirksamkeit von Arnica nicht belegt
Definierte Homöopathika/ähnliche Mittel bei definierten Indikationen						
Taylor 2000*** (448)	Allergische Erkrank.	Isopathische Nosoden/ Placebo	4 RCT	n-n-n-j-j	Meta-Analyse der Ergebnisse zu 100 mm Analogskala 9,8 (95%KI 4,2-15,4) mm besser unter Verum	Isopathische Nosoden Placebo überlegen
Vickers 2000 (476)	Erkältung	Oscillococtinum/ Placebo	7 RCT	j-j-j-j-j	Keine Überlegenheit vs. Placebo präventiv (3 Studien), jedoch Verkürzung der Krankheitsdauer kurativ	Oscillococtinum scheint die Krankheitsdauer zu verkürzen – weitere Studien notwendig
Ernst (112)	99 Chronische Kopfschmerzen	Individuelles Simile/ Placebo	4 RCT	j-t-j-j-n	1 Studie positiv, 1 mit Trend, 2 negativ	Wirksamkeit bei Behandlung chronischer Kopfschmerzen wenig wahrscheinlich
Wiesenauer 96*** (300;508)	Heuschnupfen	Galphimia/ Placebo	8 RCT, 1 KS, 2 UKS	t-n-n-j-j	Responder RR Galphimia vs. Placebo 1,25 (95%KI 1,09-1,43)	Galphimia signifikant wirksamer als Placebo

Legende siehe Tabelle 17

***Meta-Analysen ausschließlich von Studien, die die Reviewautoren selbst durchgeführt hatten

4.3.4 Merkmale und Qualitätsbewertung der Übersichtsarbeiten zu allen drei Therapien

In allen drei Therapien ist in den letzten Jahren ein starkes Ansteigen der Publikation systematischer Übersichtsarbeiten zu verzeichnen (siehe Abbildung 13). 81% der Reviews wurden als Zeitschriftenartikel publiziert, 13% elektronisch als Cochrane-Reviews (siehe Tabelle 31). Die übrigen Reviews liegen als Dissertationen oder Internetpublikationen vor. Für fast ein Drittel aller Reviews in allen drei Therapien ist eine Gruppe vom Department of Complementary Medicine (Leiter Prof. E. Ernst) der Universität Exeter in Großbritannien verantwortlich. Ein weiterer relevanter Anteil der älteren Arbeiten stammt aus dem Department of Epidemiology and Biostatistics der Universität Maastricht (damaliger Leiter Prof. P. Knipschild). Bei den neueren Arbeiten stammt ein relevanter Teil vom Zentrum für naturheilkundliche Forschung der II. Medizinischen Klinik der Technischen Universität München, dem auch der Autor dieser Schrift angehört. 22% der Reviews sind über Medline nicht auffindbar; dieser Anteil ist bei der Homöopathie mit 39% besonders hoch.

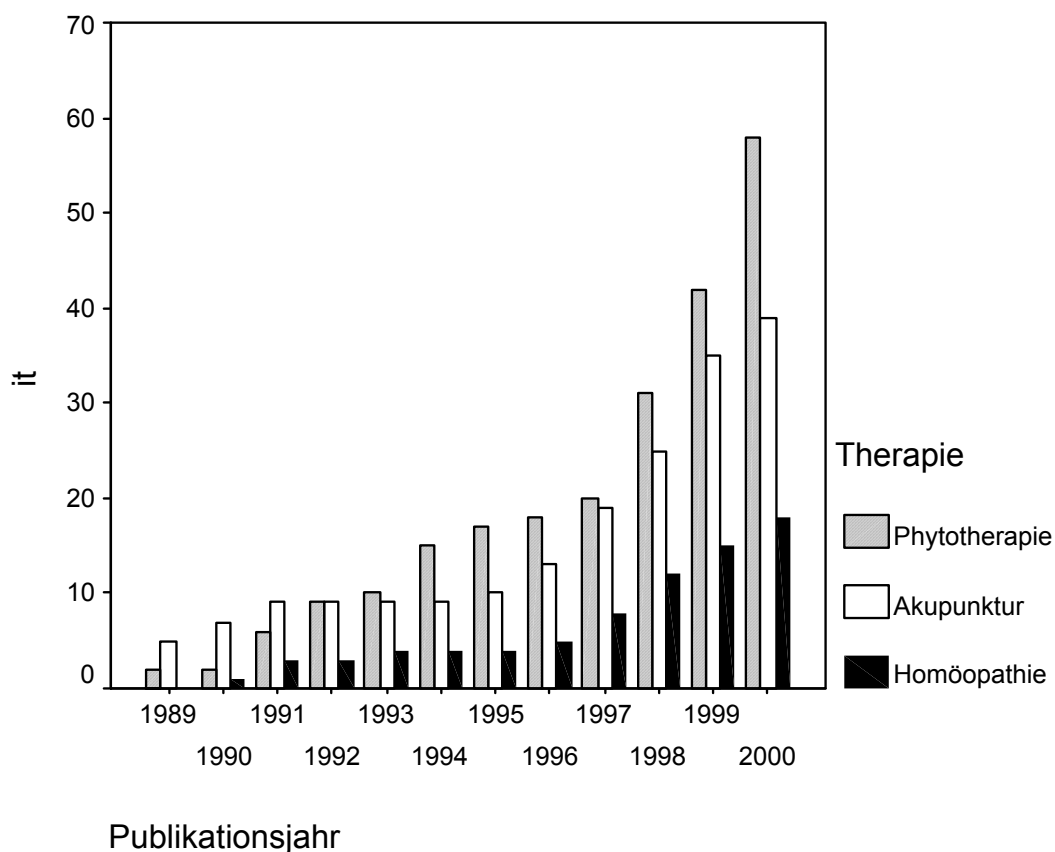


Abbildung 13:
Kumulative Zahl verfügbarer systematischer Übersichtsarbeiten zu Akupunktur, Phytotherapie und Homöopathie

In Hinsicht auf die Fragestellungen gab es zwischen den drei Therapien deutliche Unterschiede. Wie zu erwarten gab es bei den Akupunkturreviews nur selten (4% = 1 Review) eine Einschränkung auf eine klar definierte Prüfintervention (z.B. eine bestimmte Punktkombination)

.Dagegen gab es bei den Homöopathiereviews in der Regel (62%) keine Einschränkung auf eine Indikation. Im Vordergrund stand bei der Homöopathie auch besonders häufig der Vergleich mit Placebo (94%). Auffällig ist, dass der Vergleich mit anderen Interventionen in allen drei Bereichen eher sekundär ist.

Tabelle 31:
Systematische Übersichtsarbeiten zu Akupunktur, Phytotherapie und Homöopathie: Bibliographische Merkmale und Fragestellungen

	Akupunktur	Phytotherapie	Homöopathie	Insgesamt
Anzahl Reviews	39	58	18	115
- Publikationsjahr (19..)	97 (89-99)	98 (89-2000)	97 (89-99)	98 (89-2000)
- davon Cochrane Reviews	4 (10%)	9 (16%)	2 (11%)	15 (13%)
- publiziert als Artikel	34 (87%)	45 (78%)	14 (78%)	93 (81%)
- aus Exeter	11 (28%)	21 (36%)	5 (28%)	37 (32%)
- auffindbar über Medline	30 (77%)	48 (84%)	11 (61%)	89 (78%)
Fragestellung				
- eng definierte Intervention	1 (4%)	14 (37%)	3 (19%)	18 (22%)
- eng definierte Indikation	20 (74%)	33 (85%)	6 (38%)	59 (72%)
- statist. prüfbare Hypothese	12 (31%)	20 (35%)	10 (56%)	42 (37%)
- primär Vergleich mit Placebo	18 (58%)	40 (71%)	17 (94%)	75 (71%)
- primär Vergleich mit anderer Therapie	-	2 (4%)	1 (6%)	3 (3%)
- primär Vergleich mit Nichtbehandlung	2 (7%)	1 (2%)	-	3 (3%)
- gleichwertig mit Placebo/anderer Ther.?	9 (29%)	12 (21%)	-	21 (20%)
- andere/keine Vergleiche	2 (7%)	1 (2%)	-	3 (3%)

Bei vielen Reviews war die explizite Beschreibung der Einschlusskriterien nicht vollständig (siehe Tabelle 32). Dabei wurde auch eine Aussage darüber, dass auf der jeweiligen Ebene keine Beschränkung (z.B. alle Erkrankungen erlaubt) erfolgte, als explizite Beschreibung gewertet. Bzgl. Erkrankung, Intervention, Kontrollintervention und Zielkriterien fehlten Angaben in 35%, 24%, 36% und 56% der Übersichtsarbeiten. Nur bzgl. der Kontrollinterventionen ergab sich dabei zwischen den Therapien ein relevanter Unterschied: Homöopathie-Reviews machten hier häufiger klare Angaben. Ob die Publikationssprache ein Einschlusskriterium darstellte oder nicht, war in 51% der Reviews angegeben.

Entsprechend der Fokussierung auf einen Vergleich mit Placebo schlossen Homöopathie-Reviews besonders häufig nur placebokontrollierte Studien ein (44% vs. 31% bei Phytotherapie und 8% bei Akupunktur). Eine explizite Begrenzung auf randomisierte Studien erfolgte nur in 57% aller Übersichten. Insbesondere bei älteren Reviews wurden „kontrollierte klinische Studien“ eingeschlossen, wobei der Zuteilungsmodus nicht spezifiziert war. In einigen Fällen waren auch Doppelblindbedingungen als Einschlusskriterium genannt. Nur 6% aller Reviews schlossen auf der

anderen Seite unkontrollierte Studien mit ein. Ebenfalls nur 6% waren ausschließlich auf in englischer Sprache publizierte Studien begrenzt.

Tabelle 32:
Systematische Übersichtsarbeiten zu Akupunktur, Phytotherapie und Homöopathie: Angaben zu Einschlusskriterien und Literatursuche

	Akupunktur	Phytotherapie	Homöopathie	Insgesamt
Einschlusskriterien explizit bzgl.				
- Patienten/Erkrankung	29 (74%)	37 (64%)	9 (50%)	75 (65%)
- Intervention	27 (69%)	45 (75%)	15 (83%)	87 (76%)
- Kontrollintervention	21 (54%)	37 (74%)	16 (89%)	74 (64%)
- Zielkriterien	16 (41%)	29 (50%)	6 (33%)	51 (44%)
- Sprache der Publikation	20 (51%)	31 (53%)	8 (44%)	59 (51%)
Wichtige Einschlusskriterien				
- nur placebokontrollierte Studien	3 (8%)	18 (31%)	8 (44%)	29 (25%)
- nur randomisierte Studien	21 (54%)	35 (60%)	9 (50%)	65 (57%)
- auch unkontrollierte Studien	4 (10%)	2 (3%)	1 (6%)	7 (6%)
- nur in englisch publizierte Studien	4 (10%)	2 (3%)	1 (6%)	6 (6%)
Literatursuche				
- nicht beschrieben	-	8 (14%)	2 (11%)	10 (9%)
- explizit in Medline	37 (95%)	51 (88%)	13 (72%)	101 (88%)
- explizit in Embase	25 (64%)	39 (67%)	11 (61%)	75 (65%)
- explizit in anderen Datenbanken	28 (72%)	40 (69%)	12 (67%)	80 (70%)
- expl. Screening Literaturverzeichnisse	30 (77%)	46 (79%)	13 (72%)	89 (77%)
- explizit Expertenfragen	20 (51%)	29 (50%)	9 (50%)	58 (50%)
- explizit Herstellerfragen	-	38 (66%)	6 (33%)	44 (38%)

Für die Mehrheit der Reviews erfolgte die Literatursuche in einer ganzen Reihe von Quellen. Die Autoren gaben für 88% der Übersichten an, ausdrücklich in der Datenbank Medline gesucht zu haben, 65% in Embase und 70% in anderen Datenbanken. In 77% waren die Literaturverzeichnisse gefundener Arbeiten und von Reviews nach weiteren potentiellen Primärstudien durchsucht worden, in 50% waren Experten und in 38% Hersteller kontaktiert worden. Erwartungsgemäß spielte dieser letzte Suchweg bei der Akupunktur keine Rolle (kein Review vs. 66% für Phytotherapie- und 33% für Homöopathie-Reviews).

55% der Reviews machten Angaben dazu, wieviele Reviewer bei der Extraktion und Beurteilung der Studien beteiligt waren (in 49% zwei und mehr Reviewer, in 6% nur einer, siehe Tabelle 33). In 76% der Übersichtsarbeiten wurde die Qualität der Primärstudien nach vordefinierten Methoden beurteilt. In 47% der Fälle kamen Scores und in 29% der Fälle Checklisten mit Einzelkriterien zum Einsatz (dabei wendeten 4% sowohl die eine wie auch die andere Methode an). Die mediane

Anzahl eingeschlossener Primärstudien insgesamt betrug 10 (Bereich 0-107), die mediane Zahl eingeschlossener randomisierter Studien 8 (0-89). Ein Review schloss keine einzige Studie ein. Eine quantitative Meta-Analyse mit Berechnung eines gepoolten Effektmaßes erfolgte in 40% der Übersichtsarbeiten (am seltensten bei Akupunktur mit 26%). Positive bzw. eingeschränkt positive Schlussfolgerungen waren in Phytotherapiereviews am häufigsten, bei der Akupunktur am seltensten.

Tabelle 33:
Systematische Übersichtsarbeiten zu Akupunktur, Phytotherapie und Homöopathie: Methoden, Ergebnisse und Schlussfolgerungen

	Akupunktur	Phytotherapie	Homöopathie	Insgesamt
Explizit Extraktion durch ≥ 2 Reviewer	19 (49%)	31 (53%)	6 (33%)	56 (49%)
Explizit Extraktion durch 1 Reviewer	1 (3%)	3 (5%)	3 (17%)	7 (6%)
Formale Qualitätsbeurteilung				
- keine	9 (23%)	14 (21%)	4 (22%)	27 (24%)
- Score	16 (41%)	29 (50%)	7 (39%)	52 (43%)
- Einzelkriterien	11 (28%)	13 (22%)	5 (28%)	29 (25%)
- Score + Einzelitems	3 (8%)	1 (2%)	1 (6%)	5 (4%)
- andere Vorgehensweisen	-	1 (2%)	1 (6%)	2 (2%)
N eingeschl. Studien (Median, Bereich)	13 (2-51)	9,5 (0-53)	9,5 (3-107)	10 (0-107)
N eingeschlossene RCTs	8 (0-51)	7,5 (0-45)	7,5 (0-89)	8 (0-89)
Effektmaßberechnung				
- für Einzelstudien	11 (28%)	26 (46%)	11 (61%)	48 (42%)
- für Pooling	10 (26%)	26 (46%)	9 (50%)	45 (40%)
Hauptschlussfolgerung				
- eindeutig positiv	-	10 (18%)	2 (12%)	12 (11%)
- positiv mit Einschränkungen	14 (39%)	36 (63%)	7 (41%)	57 (52%)
- ambivalent	14 (39%)	10 (18%)	4 (24%)	28 (26%)
- negativ mit Einschränkungen	5 (14%)	1 (2%)	4 (24%)	10 (9%)
- eindeutig negativ	3 (8%)	-	-	3 (3%)
- nicht interpretierbar	3 (8%)	1 (2%)	1 (6%)	5 (4%)

Die Qualität der systematischen Übersichtsarbeiten war in allen drei Bereichen sehr variabel (siehe Tabelle 34). Die Beschreibung des Selektionsprozesses war nur in 38% aller Reviews ausreichend. Die meta-analytische Auswertung bzw. die Darstellung der Studienergebnisse war nur in 44% befriedigend. Bei dieser Frage sowie bei den Fragen zur Beschreibung der Auswertungsmethoden und bzgl. der Schlussfolgerungen gab es jedoch zwischen den Beurteilern vergleichsweise häufig Nichtübereinstimmungen (siehe letzte Spalte in Tabelle 34). Im Mittel wurden 6,2 von 9 möglichen Punkten bei den ersten 9 Fragen der Oxmanskala erreicht. Bei der zehnten Frage (Gesamtbeurteilung von 1 = erheblich Schwächen bis 7 = keine Schwächen) wurden im Mittel 4,6 Punkte vergeben.

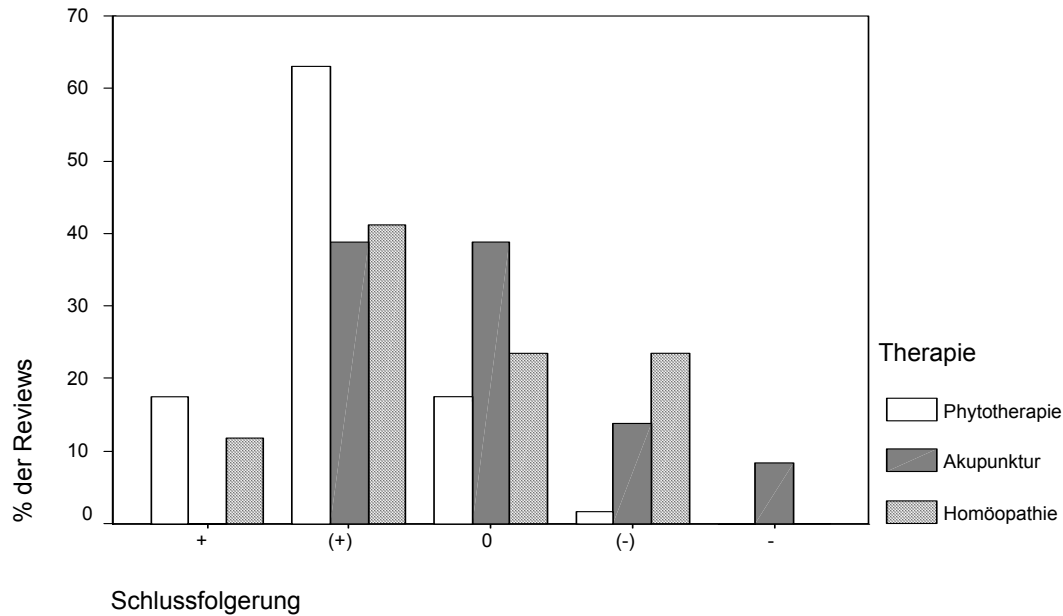


Abbildung 14:
Schlussfolgerungen der Reviewer in den Übersichtsarbeiten zu Akupunktur, Phytotherapie und Homöopathie bzgl. der vorliegenden Evidenz. + = positiv, (+) = positiv mit Vorbehalten, 0 = ambivalent, (-) = negativ mit Vorbehalten, - = negativ

Tabelle 34:
Systematische Übersichtsarbeiten zu Akupunktur, Phytotherapie und Homöopathie: Methoden, Ergebnisse und Schlussfolgerungen

Item	Akupunktur	Phytotherapie	Homöopathie	Insgesamt	Übereinst.*
Literatursuche gut beschrieben	38 (97%)	44 (76%)	12 (67%)	94 (82%)	0,66
Literatursuche ausreichend	30 (77%)	44 (75%)	13 (72%)	87 (76%)	0,69
Einschlusskriterien klar definiert	23 (59%)	37 (64%)	10 (56%)	70 (61%)	0,53
Selektionsbias vermieden	18 (46%)	22 (38%)	4 (22%)	44 (38%)	0,46
Qualitätskriterien beschrieben	28 (72%)	40 (69%)	13 (72%)	81 (70%)	0,60
Qualitätskriterien angemessen	30 (77%)	45 (78%)	14 (78%)	89 (77%)	0,56
Auswertungsmethoden beschrieben*	18 (50%)	29 (56%)	11 (61%)	58 (55%)	0,36
Angemessene Ergebnisintegration*	13 (38%)	23 (44%)	9 (53%)	45 (44%)	0,37
Schlussfolgerungen nachvollziehbar*	26 (68%)	50 (88%)	12 (71%)	88 (79%)	0,18
Gesamtbeurteilung	4,6	4,7	4,6	4,6	0,27
- Mittelwert	1,5	1,5	1,5	1,5	
- Standardabweichung	39	58	18	115	
- vorliegende Beurteilungen					
Summe der Einzelitems	6,4	6,2	6,0	6,2	
- Mittelwert	2,2	2,4	2,4	2,3	0,71
- Standardabweichung	33	52	16	101	
- vorliegende Beurteilungen					

Fehlende Daten aufgrund von Nichtbeurteilbarkeit bzw. Nichtanwendbarkeit des Kriteriums

*Übereinstimmung zwischen den Beurteilern: Für Einzelkriterien Kappa (Werte > 0,75 werden als sehr gute Übereinstimmung gewertet, Werte zwischen 0,50 und 0,74 als gute Übereinstimmung, Werte zwischen 0,25 und 0,49 als mäßige und niedrigere Werte als schlechte Übereinstimmung gewertet), für Gesamtbeurteilung und Summe Spearman Korrelationskoeffizient.

4.3.5 Vergleich der Übersichtsarbeiten zu gleichen Fragestellungen

Insgesamt wurden 21 *ähnliche* Fragestellungen (ohne Berücksichtigung von Unterfragestellungen aufgrund mehrerer Vergleiche innerhalb eines Reviews z.B. vs. Placebo und Standardtherapie) von zwei oder mehr Übersichtsarbeiten bearbeitet. 10 Fragestellungen betrafen die Akupunktur, 8

Tabelle 35:

Fragestellungen in den Bereichen Akupunktur, Phytotherapie und Homöopathie, die in mehr als einer systematischen Übersichtsarbeit untersucht wurden

Intervention und Indikation	Vergleich in allen Reviews mit	Eingeschlossene Reviews (Publikationszeitraum)	Ausgeschlossene Reviews mit Gründen
Akupunktur			
Akupunktur bei chronischen Schmerzen	Scheinakupunktur, andere Behandlung	(361;455) (1989-1990)	(131) – publiziert 2000
Akupunktur bei Nackenschmerzen	Scheinakupunktur, keine Behandlung	(415;502) (1999-2000)	-
Akupunktur bei LWS-Schmerzen	Scheinakupunktur, keine und andere Beh.	(127;415;465;466) (1998-2000)	(293) – nur Rückenschmerzen mit Ausstrahlung
Akupunktur bei Spannungskopfschmerz	Scheinakupunktur, andere Behandlung	(312;390;470) (1999-2000)	(450) – publiziert 1989
Akupunktur zur Migräneprophylaxe	Scheinakupunktur, keine und andere Beh.	(151;390) (1999)	(450) – publiziert 1989
Akupunktur zur Raucherentwöhnung	Scheinakupunktur	-	(266;503;505) – entsprechen Einschlusskriterien, jedoch Referenzen 1 und 2 zum gleichen Review und Referenz 3 vier Jahre vorher publiziert (504) – anderer Reviewansatz der Arbeitsgruppe, die Ref. 1 und 2 publizierte (454;456) – publiziert 1989 und 1990 von der gleichen Arbeitsgruppe
P6-Stimulation zur Prävention der postoperativen Übelkeit	Scheinakupunktur, keine und andere Beh.	(270;474) (1996-1999)	-
P6-Stimulation bei Schwangerschaftserbrechen	Scheinakupunktur, keine Behandlung	(7;214;474) (1996-1999)	-
Akupunktur bei Asthma	Scheinakupunktur, keine und andere Beh.	-	(246) – publiziert 1991 (290) – publiziert 1996 (280) – gleiche Arbeitsgruppe wie Referenz 2 mit eingeschr. Fragestellung
Akupunktur bei Tinnitus	Scheinakupunktur, andere Behandlung	(93;110) (1999-2000)	-

Tabelle 35 (Fortsetzung)

Fragestellungen in den Bereichen Akupunktur, Phytotherapie und Homöopathie, die in mehr als einer systematischen Übersichtsarbeit untersucht wurden

Intervention und Indikation	Vergleich in allen Reviews mit	Eingeschlossen (Publikationszeitraum)	Ausgeschlossen mit Gründen
Phytotherapie			
Ginkgo bei Claudicatio intermittens	Placebo	(330;376) (2000)	(106) – publiziert 1996 (240) – publiziert 1991 (277;401;496) – Reviews zu einem definierten Extrakt, Literatursuchen nicht beschrieben
Ginkgo bei demenziellen Erkrankungen	Placebo	-	(122) – Demenzen allgemein, publ. 1999 (349) – Begrenzung auf M. Alzheimer (191;496) – Literatursuche nicht beschrieben (241) – Demenzen publ. 1992
Hypericumextrakte bei Depression	Placebo, Standardantidepressiva	(142;235;285;483;509) (1997-2000)	(441) – nur neue Studiengruppe (104;286) – publiziert 1996 (485) – Art des Vergleichs (139) – Fokus auf Angstsy. bei Depression
Knoblauch zu Cholesterinsenkung	Placebo	Set 1: (268;443) (2000) Set 2: (411;491) (1993-1994)	(342) – Update von (411) (239;244) – allgemein zu kardiovaskulären Risikofaktoren
Echinacea bei Infekten der oberen Luftwege	Placebo	(29;320) (1999)	(322) – zu Immunmodulation allgemein
Misteltherapie bei Krebs	Placebo, keine unkonvent. Behandlung	-	(233) – Literatursuche nicht beschrieben (243) – entspricht als einziger Einschlusskriterien
Pfefferminzöl bei Colon irritabile	Placebo	(207;373) (1998-2000)	-
Sabal bei benigner Prostatahyperplasie	Placebo, Standardmedikation	-	(49) – beschränkt auf ein Extrakt (511) - zu Sabalprodukten allgemein
Homöopathie			
Homöopathie bei verschiedenen Erkrankungen	Placebo, Behandlungen	andere (185;245) (1990-1991)	-
Homöopathie bei verschiedenen Erkrankungen	Placebo	(76;278;489) (1997-2000)	-
Arnica bei verschiedenen Indikationen (meist Trauma)	Placebo	(119;301)	-

die Phytotherapie und 3 die Homöopathie (siehe Tabelle 35). Unter Anwendung der engergefassten Einschlusskriterien bzgl. *gleicher* Fragestellungen und Reviewqualität blieben 16 Fragestellungen (8 zur Akupunktur, 5 zur Phytotherapie und 3 zur Homöopathie) für die

vergleichende Analyse. Bei einer Fragestellung (Cholesterinsenkung durch Knoblauchpräparate) mußten aufgrund des unterschiedlichen Alters der Reviews zwei separate Vergleiche bearbeitet werden (zwei Reviews, die im Jahr 2000 publiziert wurden sowie zwei Reviews, die 1993 und 1994 publiziert wurden). Dementsprechend erfolgten insgesamt 17 Vergleiche.

Bei 15 von 17 Vergleichen unterschied sich die Zahl bzw. die Gruppe der eingeschlossenen Primärstudien unter Berücksichtigung des Publikationsjahres der Reviews relevant (d.h. um mehr als 25%; siehe Tabelle 36). Bei 10 Vergleichen variierte die Primärstudienzahl um mehr als 50%. Häufigster Grund für die Diskrepanzen waren Unterschiede bei den Einschlusskriterien und in der Literatursuche.

Tabelle 36:

Zahl der in die Übersichtsarbeiten zu klinischen Studien zu Akupunktur, Phytotherapie und Homöopathie eingeschlossenen Primärstudien und Gründe für Diskrepanzen

Intervention und Indikation	Zahl der Primärstudien* (Gesamtzahl identifizierter Studien)	Hauptgründe für relevante Diskrepanzen
Akupunktur		
Akupunktur bei chronischen Schmerzen	46 (455), 14 (361) (48)	Literatursuche, Einschlusskriterien
Akupunktur bei Nackenschmerzen	5 (415), 11 (502) (11)	Vermutlich Literatursuche, weniger Einschlusskriterien
Akupunktur bei LWS-Schmerzen	8 (415), 11 (465), 10 (127) (11)	Einschlusskriterien und Literatursuche
Akupunktur bei Spannungskopfschmerz	6 (312), 8 (390), 7 (470) (9)	Einschlusskriterien
Akupunktur bei Migräne	16 (390), 6 (151) (16)	Wahrscheinl. Berücksichtigung bzw. Nichtberücksichtigung von nicht-englischsprachigen Publikationen
P6-Stimulation und postop. Übelkeit	19 (270), 18 (474) (24)	Alter (1996 vs. 1999) und Einschlusskriterien
P6-Stimulation bei Schwangerschaftsübelkeit	7 (7), 4 (214), 7 (474) (8)	Einschlusskriterien
Akupunktur bei Tinnitus	6 (360), 6 (93) (7)	
Phytotherapie		
Ginkgo bei Claudicatio intermittens	8 (376), 5 (330) (9)	Einschlusskriterien und Literatursuche
Hypericum bei Depression	8 (142), 14 (509), 6 (235), 27 (285), 11 (483) (29)	Primär Einschlusskriterien
Knoblauch zur Cholesterinsenkung - Reviews 2000 - Reviews 1993-1994	37 (268), 13 (443) (37) 13 (411), 5 (491) (13)	Einschlusskriterien und Literatursuche Einschlusskriterien und Literatursuche
Echinacea bei Infekten der ob. Luftwege	13 (29), 13 (320) (15)	
Pfefferminzöl bei Colon irritabile	3 (207), 7 (373) (8)	Vermutlich primär Suche, evtl. Selektionsprozess
Homöopathie		
Homöopathie bei verschiedenen Erkrankungen vs. Placebo	16 (76), 36 (489), 89 (278) (97)	Fast ausschliesslich Einschlusskriterien
Homöopathie bei verschiedenen Erkrankungen	105 (245), 40 (185) (105)	Einschlusskriterien und Literatursuche
Arnica bei verschiedenen Erkrankungen	35 (300), 8 (111) (35)	Primär Einschlusskriterien, Selektionsprozess

*Neuere Reviews zuerst

Ein erstes fundamentales Problem beim Vergleich der vorliegenden Übersichtsarbeiten war, dass die Hauptfragestellungen sehr häufig eher global-deskriptiv und nur selten genauer operationalisiert waren. Box 1 gibt ein typisches Beispiel.

Box 1

Fragestellungen bzw. Ziele in systematischen Übersichtsarbeiten zu Echinaceaextrakten

Barrett et al. (29): "To review the evidence regarding the effectiveness of orally ingested Echinacea extracts in reducing the incidence, severity, or duration of acute upper respiratory infections."

Melchart et al. (320): "To assess whether there is evidence that preparations containing extracts of Echinacea are effective in the prevention and treatment of the common cold."

Ohne eine Überprüfung der Einschlusskriterien bzw. der tatsächlich eingeschlossenen Studien lässt sich daher in der Regel nicht sicher sagen, ob die Fragestellungen wirklich identisch sind.

Dementsprechend kommt einer genauen Prüfung der Einschlusskriterien und des Selektionsprozesses entscheidende Bedeutung zu. In den vorliegenden systematischen Übersichtsarbeiten ergab die detaillierte Prüfung der Einschlusskriterien, dass in der Mehrzahl der Fälle Unterschiede bestanden, die zwar oft nur wenige Worte betrafen, jedoch zum Teil dazu führten, dass sich die Zahl der eingeschlossenen Studien dramatisch unterschied. In Tabelle 37 sind als Beispiel die Einschlusskriterien der zwischen 1997 und 2000 publizierten systematischen Übersichtsarbeiten zur Depressionsbehandlung mit Hypericumextrakten – soweit möglich im Originalwortlaut – und die Anzahl der jeweils eingeschlossenen placebokontrollierten Studien wiedergegeben. Die Zahl der eingeschlossenen Primärstudien variiert zwischen 2 und 17 bei insgesamt 20 Studien, die in mindestens einem Review berücksichtigt waren. Die unterschiedliche Verfügbarkeit aufgrund des Alters der Reviews spielte dabei eine untergeordnete Rolle. Die Einschlusskriterien, die die Studienzahl am meisten einschränkten, sind in Tabelle 37 in kursiver Schrift dargestellt. So ist der primäre Grund der geringen Studienzahl im Review von Gaster (142) die Begrenzung auf Publikationen in englischer Sprache. Bei Kim et al. (235) ist die geringe Studienzahl praktisch ausschliesslich durch die Begrenzung auf Untersuchungen, in denen die Klassifikation der Depression explizit nach ICD 10 oder DSM Kriterien erfolgte, begründet. Bei Williams et al. (509) ist die Mindestdauer der Behandlung von 6 Wochen das entscheidende Kriterium. Eine Nichtidentifikation einer relevanten Zahl von Studien kann in den neueren Reviews praktisch ausgeschlossen werden, da diese einen älteren und bereits deutlich umfangreicheren Review (286) zitieren.

Dagegen hatte z.B. ein potentiell sehr relevanter Unterschied in den Einschlusskriterien im Fall von Hypericum faktisch nur einen minimalen Einfluss auf die Primärstudienzahl: Der Einbezug bzw. Nichteinbezug von Kombinationspräparaten (d.h. Präparaten, in denen neben Hypericum auch andere Pflanzenextrakte enthalten sind). Im Gegensatz hierzu würde z.B. bei

Echinaceapräparaten der Ausschluss von Kombinationspräparaten die Studienzahl erheblich reduzieren.

Tabelle 37:
Einschlusskriterien in systematischen Übersichtsarbeiten zu Hypericumextrakten bei Depression und Anzahl der jeweils eingeschlossenen placebokontrollierten Studien

Review	Patienten bzw. Indikation	Intervention	Design	Andere	Studien
Volz (485)	Depressed patients	Hypericum mono-preparations	Double-blind clinical trial	-	11
Linde (285)	Depressive disorders	1. Hypericum monoprep. 2. Combinations	Randomized clinical trials	-	16 1
Kim (235)	<i>Depressive disorder acc. to ICD10/DSM</i>	Hypericum mono-preparations	Blinded controlled trials	Outcome Hamilton Rating Scale Depression	2
Gaster (142)	Depressive disorder	Hypericum mono-preparations	Double-blind RCTs	<i>Available in English</i>	4
Williams (509)	Depressive disorders	Herbal preparations incl. Hypericum	RCTs	<i>At least 6 weeks duration</i>	8

Weitere Beispiele von dramatischen Unterschieden in der Primärstudienzahl aufgrund von divergierenden Einschlusskriterien sind die zwei neueren Reviews zur Cholesterinsenkung mit Knoblauchpräparaten (Unterschiede bzgl. Blindbedingungen und präsentierten Daten; zusätzlich wahrscheinlich Unterschiede bei der Literatursuche) sowie die drei Meta-Analysen zur Frage, ob Homöopathie Effekte über Placebo hinaus hat (Begrenzung auf Studien mit vordefiniertem Hauptzielkriterium in (76) und auf in den Datenbanken Medline oder Embase gelisteten Studien in (489) gegenüber Einschluss aller Studien mit verwertbaren Ergebnisdaten in (278))

Aufgrund der Komplexität der Literatursuchen und den daraus resultierenden Ungenauigkeiten in der Beschreibung ist die Vergleichbarkeit der Suchmethoden und deren Auswirkungen auf die Vollständigkeit selbst für methodisch erfahrene Leser schwer zu beurteilen. Während für Suchen in der Datenbank Medline zumindest zum Teil Details wie Suchbegriffe und Suchzeitraum angegeben sind, die eine Reproduzierbarkeit bis zu einem gewissen Grade gewährleisten, ist dies für andere Datenbanken fast nie der Fall. Die Effizienz anderer Suchmethoden wie Kontakte mit Experten oder das Durchsuchen der Literaturangaben gefundener Arbeiten ist ebenfalls nicht überprüfbar. In den vorliegenden Übersichtsarbeiten waren die verwendeten Suchstrategien in keinem einzigen Fall genau identisch. Die Auswirkungen der Unterschiede bei den Suchstrategien sind aber nur durch den Vergleich der jeweils tatsächlich zitierten Studien unter Berücksichtigung der Einschlusskriterien abschätzbar. Als Beispiel sind in Box 2 die Suchmethoden der beiden Übersichtsarbeiten zu Echinacea vereinfacht zusammengefasst.

Box 2

Literatursuchmethoden bzw. -quellen in systematischen Übersichtsarbeiten zu Echinaceaextrakten

Barrett et al. (29): Medline (Suchbegriff Echinacea), Kontakt mit Experten, Durchsuchen der Literaturangaben gefundener Arbeiten

Melchart et al. (320): Datenbanken Medline, Embase, Cochrane Acute Respiratory Infections Collaborative Review Group,

Cochrane Complementary Medicine Field, Phytodok mit dem Suchbegriff Echinac*; Durchsuchen der Literaturangaben gefundener Arbeiten, Kontakt mit Experten und Herstellern, Abstractbänden von Kongressen

Obwohl die Suchmethoden in der Übersicht von Melchart et al. (320) wesentlich umfangreicher sind, ist die Studiensammlung von Barrett et al. (29) fast identisch (vermutlich weil Barrett viele Studien mit Hilfe eines früheren Review von Melchart et al. (322) erfasste, der bereits die meisten Studien eingeschlossen hatte). Dagegen können auf den ersten Blick ähnlich umfangreiche Literatursuchen wie im Beispiel der beiden Knoblauch-Reviews (268;443) offensichtlich deutlich unterschiedliche Suchergebnisse erbringen.

Die Beschreibung des Selektionsprozesses (der Prozess, in dem aus den durch die Literatursuche identifizierten Referenzen die letztlich eingeschlossenen Primärstudien selektiert werden) ist in vielen Übersichtsarbeiten nicht genau genug, um mit Sicherheit ausschliessen zu können, dass trotz gleicher Einschlusskriterien unterschiedliche Studien ausgewählt werden. In einigen Fällen wurden nicht nachvollziehbare Diskrepanzen festgestellt, z.B. bei den Übersichtsarbeiten zu Akupunktur bei Nackenschmerzen (415;502). Hier ist anzunehmen, dass die als Einschlusskriterium genannte Indikation Nackenschmerzen unterschiedlich definiert wurde, sodass Smith et al. (415) Studien ausschlossen, die White und Ernst (502) berücksichtigten.

Aufgrund der häufig unterschiedlichen Instrumente zur Qualitätsbeurteilung ist ein unmittelbarer Vergleich der Bewertungen der einzelnen Primärstudien schwierig. Der Großteil aller Reviews bemängelt aber die überwiegend mäßige Qualität der Primärstudien. Deutliche Diskrepanzen in der Gesamtbeurteilung traten eigentlich nur bei den Übersichtsarbeiten zur Akupunktur bei Schmerzen im Bereich der Lendenwirbelsäule auf. Ernst und White bewerteten die Qualität als überwiegend gut („good in the majority of studies“ (127), S. 2237), van Tulder et al. als extrem schlecht („methodological quality was ... extremely poor“ (465), S. 1120) und Smith et al. als überwiegend schlecht („majority of trials were of poor quality“ (415), S. 123).

Die Methoden zur Zusammenfassung der Ergebnisse unterschieden sich in mehr als der Hälfte der Vergleiche in relevanter oder potentiell relevanter Weise. Zum Einsatz kamen deskriptive Ergebniszusammenfassungen, Ergebniskategorisierungen (vote counts) und Effektstärkeberechnungen mit und ohne Meta-Analysen. Wiederum waren hier die Diskrepanzen der drei Übersichtsarbeiten mit Studien zur Akupunktur bei Schmerzen im Bereich der Lendenwirbelsäule besonders auffällig. Ernst und White (127) führten quantitative Effektstärkeberechnungen einschl. Meta-Analyse durch, van Tulder et al. (465;466) Effektstärkeberechnungen und Ergebniskategorisierungen und Smith et al. (415) Ergebniskategorisierungen. Die beiden letzten

Reviews präsentierten außerdem deskriptive Zusammenfassungen der Studienergebnisse. Die Ergebniskategorisierungen von van Tulder et al. und Smith et al. unterscheiden sich sowohl bzgl. der Kategoriendefinition als auch bzgl. der letztlich erfolgten Beurteilungen (siehe Tabelle 38). In beiden Reviews wird zwischen den Interpretationen der Primärstudienautoren und der Reviewer unterschieden, aber sogar hinsichtlich der ersteren gibt es relevante Differenzen.

Tabelle 38:
Ergebniskategorisierungen der Rückenschmerzstudien, die sowohl in den Review von van Tulder et al. (465;466) als auch in denjenigen von Smith et al. (415) eingeschlossen worden waren. Van Tulder et al. wendeten eine dreistufige Kategorisierung an (positiv, neutral, negativ) mit einer Option unklar, während Smith et al. die Studien als positiv oder negativ bewerteten. Beide Reviewteams führten separate Kategorisierungen der Schlussfolgerungen der Autoren der Primärstudien und der Reviewer durch.

	van Tulder et al.	Smith et al.	van Tulder et al.	Smith et al.
Studie	Schlussfolgerung Autor	Schlussfolgerung Autor	Schlussfolg. Reviewer	Schlussfolg. Reviewer
Coan et al.	Positiv	Positiv	Unklar	Positiv
Duplan et al.	Positiv	Positiv	Positiv	Negativ
Edelist et al.	Neutral	Negativ	Neutral	Negativ
Garvey et al.	Positiv	Negativ	Neutral	Negativ
Gunn et al.	Positiv	Positiv	Neutral	Positiv
Lehmann et al.	Positiv	Negativ	Neutral	Negativ
Lopacz & Gralewski	Neutral	Negativ	Neutral	Positiv
Mendelson et al.	Neutral	Negativ	Unklar	Negativ

Die Verwendung unterschiedlicher meta-analytischer Methoden ist am deutlichsten am Beispiel der Übersichtsarbeiten zu dem Thema, ob Homöopathie Effekte über Placebo hinaus bewirkt. Cucherat et al. (76) verwendeten eine Methode, bei der die p-Werte der einzelnen Studien kombiniert wurden, Walach (489) berechnete standardisierte Mittelwertsdifferenzen und Linde et al. (278) Odds Ratios. Dies sind völlig unterschiedliche Ansätze zur Quantifizierung von Effekten. In allen drei Meta-Analysen wurden dagegen übereinstimmend für die primären Analysen sogenannte random effects Modelle verwendet (d.h. es wird angenommen, dass der „wahre“ Effekt in den einzelnen Studien differiert. Bei fixed effects Modellen wird dagegen angenommen, dass der „wahre“ Effekt in allen Studien gleich ist). Bei zwei der drei Meta-Analysen (76;278) ergab sich eine signifikante Überlegenheit der homöopathischen Interventionen gegenüber Placebo, bei der dritten (489) dagegen nicht. Dies ist vermutlich auf die Wahl eines speziellen random effects Modells (Hunter und Schmidt) zurückzuführen, welches bei großer Zwischenstudienheterogenität zu extrem weiten Konfidenzintervallen führt.

Eine Auswirkung des gewählten Meta-Analyse-Modells war jedoch nur bei dieser Fragestellung zu erkennen. In den drei anderen Fällen, in denen in mehr als in einem Review eine Meta-Analyse

durchgeführt worden war (Hypericum bei Depression und bei beiden Reviewsets zu Knoblauch zur Cholesterinsenkung), waren die (geringen) Unterschiede der berechneten Effektmaße mit großer Wahrscheinlichkeit auf die Unterschiede hinsichtlich der einbezogenen Primärstudien zurückzuführen.

Im Vergleich zu den großen Diskrepanzen bei Einschlusskriterien und Primärstudienzahl sind die Differenzen bei den Schlussfolgerungen gering. Die Schlussfolgerungen sind aber auch in den meisten Fällen angesichts der Evidenzlage sehr zurückhaltend. Differenzen scheinen eher durch die Unterschiede in der grundsätzlichen Haltung der Autoren zu der Intervention begründet als in den Daten selbst.

4.4 Diskussion

4.4.1 Zusammenstellung der systematischen Übersichtsarbeiten zu Akupunktur, Phytotherapie und Homöopathie

Die Zusammenstellung zeigt, dass zu den untersuchten komplementärmedizinischen Verfahren eine nicht unerhebliche Anzahl von systematischen Übersichtsarbeiten vorliegt. Dennoch ist die Evidenz nur in Einzelfällen eindeutig.

Bei der Akupunktur wurden nur im Falle der P6-Stimulation zur Prävention der postoperativen Übelkeit (Wirksamkeit im Vergleich zur Scheinakupunktur belegt und ähnliche Wirksamkeit wie gängige Antiemetika) und der Raucherentwöhnung (kein Effekt einer Verum- im Vergleich zu einer Scheinakupunktur) eindeutige Schlussfolgerungen gezogen. Angesichts der weiten Verbreitung der Akupunktur und der vergleichsweise großen Zahl von Primärstudien ist diese Tatsache frustrierend. Das Hauptproblem in den Akupunkturstudien scheinen die unzureichenden Fallzahlen zu sein. Die medianen Fallzahlen der Studien zu Rückenschmerzen, Kopfschmerzen und Schmerzen bei Osteoarthritis waren z.B. 50 (465), 37 (390) und 31 (107). Ein zweites großes Problem ist die methodische Qualität der Primärstudien. Praktisch alle Reviews enden mit dem Ruf nach größeren und methodisch besseren Studien.

Ein möglicher Grund für das Fehlen solcher Studien dürfte – zumindest in der Vergangenheit – die fehlende finanzielle Förderung derartiger Studien gewesen sein. Das Industrieinteresse an Akupunktur ist gering, sodass eine relevante Förderung nur durch die öffentliche Hand realistisch erscheint. Eine solche Förderung hat erst in den letzten Jahren eingesetzt. In Deutschland, den USA und Grossbritannien sind derzeit mehrere große Studien in der Planung, in Durchführung oder zum Teil bereits abgeschlossen und im Publikationsprozess. Ein weiterer und verwandter Grund war bislang das Fehlen einer adäquaten Forschungsinfrastruktur, d.h. von

Personengruppen und Einrichtungen, die über ausreichendes Know-how sowohl in der Akupunktur wie auch in der klinischen Forschung verfügten. Ein dritter Grund ist, dass klinische Forschung zur Akupunktur in der Tat methodisch schwierig ist. Der Begriff Akupunktur bezeichnet (vergleichbar mit der Situation bei anderen nichtmedikamentösen Verfahren wie z.B. der Physiotherapie) eine heterogene Gruppe von Interventionen. Aufgrund von Ergebnissen zu einer definierten Akupunkturstrategie kann kaum eine Aussage zur Akupunktur insgesamt abgeleitet werden. Äußerst schwierig ist auch die Wahl einer adäquaten Kontrollgruppe. Aufgrund der aus westlicher Sicht geringen Plausibilität der Akupunktur wird häufig der Nachweis der „spezifischen“ Wirksamkeit gegenüber Placebo gefordert. Verschiedenste Vorgehensweisen wurden als „Placebo“ eingesetzt, wie z.B. die Nadelung an nicht-indizierten Punkten, an Nichtakupunkturpunkten (mit tiefer oder oberflächlicher Nadelung), ausgeschaltete Laserakupunktur- oder TENS-Geräte (transkutane elektrische Nervenstimulation), sogenannte Placebo-Nadeln (in ein Pflaster gesteckte Teleskopnadeln, die nicht in die Haut eindringen) etc. Die Verfahren, bei denen eine Hautpenetration erfolgt, sowie Verfahren, bei denen ein starker Druckreiz an Akupunkturpunkten erfolgt, können sicher nicht als physiologisch inert angesehen werden und könnten damit zu einer Unterschätzung der Wirksamkeit der Akupunktur führen. Dagegen sind z.B. bei der Verwendung eines ausgeschalteten Laserakupunkturgerätes Blindbedingungen nur möglich, wenn die Patienten nicht vollständig aufgeklärt werden. Angesichts dieser massiven methodischen Probleme sind die heterogenen und zum Teil widersprüchlichen Ergebnisse der vorliegenden Akupunkturstudien durchaus plausibel.

Mehrere der Reviews fanden eine negative Korrelation zwischen Studienqualität und Ergebnis, d.h. bessere Studien waren öfter negativ (z.B. (131;415)). Diese Resultate sind mit Zurückhaltung zu interpretieren. Die entsprechenden Reviews schlossen Studien mit sehr unterschiedlichen Kontrollgruppen ein: Wartelistenkontrollen bzw. keine Behandlung, die verschiedenen beschriebenen Scheinakupunkturverfahren und verschiedene aktive Kontrollinterventionen. Es ist nicht davon auszugehen, dass die Unterschiede zwischen Akupunktur und diesen verschiedenen Kontrollen alle gleich sind. Z.B. würde man bei Studien mit Wartelistenkontrolle größere Unterschiede erwarten als bei Studien mit Scheinakupunkturkontrolle. Akupunkturstudien ohne Scheinakupunkturkontrolle können aber nicht verblindet werden und schneiden daher bei der Qualitätsbewertung typischerweise schlechter ab. Untersuchungen zu Zusammenhängen zwischen Studienqualität und Ergebnis sollten daher grundsätzlich nur bei Studien mit gleicher Kontrollintervention durchgeführt werden.

Im Bereich der Phytotherapie sind die Ergebnisse und Schlussfolgerungen der meisten Übersichtsarbeiten tendenziell positiv. Aufgrund methodischer Probleme oder geringer Primärstudienzahlen sind definitive Schlussfolgerungen aber auch hier die Ausnahme. Auffällig ist die Häufung systematischer Übersichtsarbeiten zu Hypericum, Ginkgo und Knoblauch. Dies liegt

vermutlich an der relativ großen Zahl verfügbarer Primärstudien und an der Tatsache, dass es sich um weitverbreitete und für die Pharmaindustrie relevante Phytotherapeutika handelt. Insgesamt beschränken sich die Phytotherapie-Reviews – ebenso wie die überwiegende Zahl der Primärstudien – auf Extrakte, die in der Vermarktung eine Rolle spielen und bei denen Chancen auf eine Zulassung als Arzneimittel am europäischen oder nordamerikanischen Markt bestehen. Allerdings liegen für einige weit verbreitete Extrakte (z.B. Crataegusextrakte) bisher noch keine systematischen Übersichtsarbeiten vor. Viele traditionelle phytotherapeutische Anwendungen wie Teeanwendungen oder phytotherapeutische Strategien in Entwicklungsländern werden aber aufgrund mangelnden finanziellen Interesses gar nicht untersucht.

Bei der Phytotherapie erscheint die Gefahr des publication bias (242), d.h. der Nichtveröffentlichung negativer Ergebnisse, besonders groß. Phytotherapeutika werden häufig von mittelständischen Unternehmen hergestellt, für die eine negative Studie eine existenzielle Bedrohung darstellen kann. Allerdings kommt es aufgrund der fehlenden Patentierbarkeit auch zur gegenteiligen Situation: Hersteller verhindern die Publikation positiver Studien zu ihrem Präparat, damit die Daten nicht von einem Konkurrenten für die Zulassung seines Präparates mit einem ähnlichen Extrakt verwendet werden können.

Ein weiteres großes Problem von systematischen Übersichtsarbeiten zu Pflanzenextrakten ist die häufig unklare Vergleichbarkeit unterschiedlicher Produkte bzw. Extrakte. Dieses Problem spielt in der Expertendiskussion eine große Rolle, ist aber für die meisten Leser einer Übersichtsarbeit kaum einschätzbar. Besonders verwirrend ist die Situation z.B. bei Echinaceaprodukten: Es gibt drei verschiedene, medizinisch relevante Echinaceaarten (*E. purpurea*, *pallida* und *angustifolia*), verwendet werden Wurzeln, oberirdische Anteile, nur Blätter oder verschiedene Teile, es gibt hydro- und lipophile Extraktionsmethoden und letztendlich sind in vielen marktüblichen Präparaten weitere Pflanzenextrakte beigemischt. Die Durchführung von quantitativen Meta-Analysen ist in einer solchen Situation nicht sinnvoll. Sichere Aussagen lassen sich letztlich meist nur für die einzelnen geprüften Extrakte machen.

Bei den Übersichtsarbeiten zur Homöopathie spielt die grundsätzliche Frage nach der Wirksamkeit im Vgl. zu Placebo unabhängig von der Indikation oder der speziellen homöopathischen Vorgehensweise eine dominierende Rolle. Konkrete klinische Fragestellungen wurden nur in etwa der Hälfte der Übersichten bearbeitet. Von diesen Fragestellungen ist wiederum nur ein Teil für die homöopathische Praxis relevant, während sich andere Reviews mit nur aus Forschungssicht potentiell bedeutsamen Themen befassen (z.B. der Behandlung des postoperativen Ileus oder des Muskelkaters). Dementsprechend können die vorliegenden Übersichten für klinisch tätige Homöopathen und „Schulmediziner“ keine Evidenzbasis für ärztliche Entscheidungen bilden.

Es muss nochmals darauf hingewiesen werden, dass die vorliegende Zusammenstellung der Übersichtsarbeiten zu Akupunktur, Phytotherapie und Homöopathie nur einen sehr groben Überblick der Evidenzlage zu diesen Verfahren geben kann. Es kann nicht das Ziel einer solchen Zusammenstellung sein, die Wirksamkeit dieser Verfahren differenziert zu beurteilen. Die Zusammenfassung eines systematischen Reviews, der selbst wiederum die Zusammenfassung einer Reihe von Primärstudien ist, in wenigen Sätzen oder in einer Tabelle muss Ergebnisse und Schlussfolgerungen in hohem Maße vereinfachen. Für genauere Informationen oder wenn klinische Entscheidungen getroffen werden sollen, ist es daher für den Leser unumgänglich, die jeweiligen systematischen Übersichtsarbeiten genau zu studieren. Die Zusammenstellung erleichtert jedoch den Zugang hierzu in hohem Maße und gibt eine Orientierung, zu welchen Fragestellungen überhaupt Reviews vorliegen.

4.4.2 Merkmale und Qualitätsbewertung der Übersichtsarbeiten zu allen drei Therapien

Die Analyse der Merkmale und Qualität der bearbeiteten systematischen Reviews zeigt, dass (1) sich die Art der Fragestellungen in den drei Verfahren deutlich unterscheidet, (2) viele Übersichtsarbeiten relevante methodische Defizite bzw. Defizite bei der Beschreibung der angewandten Methoden haben und dass (3) die Schlussfolgerungen zur Phytotherapie am häufigsten positiv und bei der Akupunktur besonders häufig ambivalent bzw. weniger positiv sind.

Ad (1): Die Fragestellungen der vorliegenden Übersichtsarbeiten sind am spezifischsten in der Phytotherapie, d.h. sie sind vergleichsweise klar definiert in Bezug auf die Prüf- und Kontrollinterventionen sowie auf die Indikation, z.B. „Sind Hypericumextrakte wirksamer als Placebo bei der Behandlung von Depressionen?“ Bei der Akupunktur sind in der Regel die Indikationen definiert, die Intervention ist jedoch meist die „ganze“ Akupunktur und in Bezug auf die Interventionen in der Kontrollgruppe sind die Einschlusskriterien meist weit, z.B. „Ist (jede Form der) Akupunktur wirksam bei Asthma?“ Bei der Homöopathie ist die Fragestellung häufig noch allgemeiner, z.B. „Ist Homöopathie (bei allen Indikationen in jeder Anwendungsweise) wirksamer als Placebo?“ Die Ansätze reflektieren einerseits die unterschiedliche Plausibilität der Wirkungsmechanismen sowie die Eigenheiten der Verfahren. So ist bei der Akupunktur eine Einschränkung auf definierte Einzelstrategien ähnlich wie bei anderen nichtmedikamentösen Verfahren wie z.B. der Chirotherapie oder der Physiotherapie nur in Einzelfällen möglich. Es muss aber klar sein, dass die Ergebnisse und Schlussfolgerungen aus Übersichtsarbeiten mit so unterschiedlich spezifischen Fragestellungen nicht in gleicher Weise für klinische Entscheidungen herangezogen werden können. Im Vergleich zu Übersichtsarbeiten zu konventionellen Pharmaka ist auch bei den Phytopharmaka, soweit es sich nicht um Übersichten zu einem einzigen Extrakt handelt, aufgrund der Variabilität die Fragestellung in Bezug auf die Intervention weniger spezifiziert.

Ad (2): Ähnlich wie bei den Primärstudien ist auch bei den Übersichtsarbeiten die Qualität außerordentlich heterogen. Die verwendete Beurteilungsskala von Oxman et al. (355) hat sich bei der Beurteilung nur in eingeschränktem Maße bewährt. Ein Problem mit der Skala, die ursprünglich für die Beurteilung von Übersichtsarbeiten allgemein und nicht von *systematischen* Übersichtsarbeiten im Speziellen entwickelt wurde, ist, dass nur sehr kurzgefasste Anleitungen für die Beantwortung der einzelnen Fragen vorliegen (persönliche Kontakte mit Andrew Oxman, 1997), die einen weiten Beurteilungsspielraum lassen. Für die durchgeführte Analyse wurden nur für einige besonders unklare Punkte (wie z.B. die Beurteilung der Studienselektion) Konkretisierungen in den Anleitungen vorgenommen. Weitergehende Eingriffe wurden als problematisch angesehen, da die Validierung der Skala mit den Originalanleitungen erfolgt war (356). Bei einigen Fragen (zur Beschreibung und Angemessenheit der Auswertungsmethoden sowie zur Nachvollziehbarkeit der Schlussfolgerungen) war aber nun in unserer Analyse die Übereinstimmung zwischen den Reviewern unbefriedigend. Außerdem erschien in der Zwischenzeit eine Analyse von systematischen Übersichtsarbeiten zu verschiedenen Interventionen bei Asthma von Jadad et al. (205), in der die Oxmanskala ebenfalls verwendet wurde. Die Beurteilungsanleitungen waren jedoch für diese Analyse – insbesondere für die Gesamtbewertung in Frage 10 - in hohem Maße verschärft worden. Zwei der von Jadad et al. beurteilten Reviews wurden auch in die hier vorgelegte Analyse eingeschlossen (280;281). Zwar war die Beurteilung bei den Fragen 1 bis 9 weitgehend gleich, bei Frage 10 wurden die beiden Reviews von Jadad et al. (jeweils aufgrund eines definierten Kritikpunkts) als fehlerhaft, in der vorliegenden Analyse dagegen als weitgehend fehlerfrei eingestuft. Es macht wenig Sinn, sich auf eine validierte Skala zu beziehen, wenn die Anleitungen für die Bewertung deutlich modifiziert werden. Für die Beurteilung von systematischen Übersichtsarbeiten sind die Originalanleitungen von Oxman et al. aber einfach zu ungenau. Wie bei der formalen Beurteilung von klinischen Studien gilt jedoch auch bei systematischen Übersichtsarbeiten grundsätzlich, dass mit Checklisten und Skalen nur eine grobe Bewertung möglich ist.

1999 wurden Leitlinien für die Publikation von Meta-Analysen publiziert (327), die möglicherweise zu einer Verbesserung der Qualität und der Berichterstattung führen werden. Einige der enthaltenen Empfehlungen sind jedoch nur auf quantitative Meta-Analysen mit sehr eng definierten Fragestellungen anwendbar, während sie für systematische Übersichtsarbeiten mit weiteren Fragestellungen nicht umsetzbar sind.

Ad (3): Bei den häufig positiven Schlussfolgerungen bei der Phytotherapie muss berücksichtigt werden, dass fast die Hälfte der systematischen Übersichtsarbeiten zu Ginkgo, Hypericum und Knoblauch durchgeführt wurden, bei denen die Evidenz im Vergleich zu anderen Pflanzen gut ist. Bei der Akupunktur ist die in 4.4.1. diskutierte Heterogenität von Prüf- und Kontrollinterventionen und die geringe Fallzahl zu beachten.

Dem Autor dieser Monographie sind nur wenige vergleichbare deskriptiv-epidemiologische Analysen zu Übersichtsarbeiten in der konventionellen Medizin bekannt (z.B. (204)). Solche Untersuchungen erscheinen sinnvoll, um allgemeine Trends in den Vorgehensweisen, Schwachstellen und Entwicklungen zu erfassen. Im Rahmen der Cochrane Collaboration werden aber derzeit zahlreiche empirische Untersuchungen zu verschiedensten Aspekten systematischer Reviews durchgeführt (siehe Cochrane Methodology Register in der Cochrane Library), die zu weiteren methodischen Entwicklungen führen werden.

4.4.3 Vergleich der Übersichtsarbeiten zu gleichen Fragestellungen

Trotz der zu erwartenden methodischen Weiterentwicklungen in Bezug auf systematische Übersichtsarbeiten sind die Ergebnisse der vergleichenden Analyse der Reviews zu gleichen Fragestellungen beunruhigend. Es wurde gezeigt, dass Einschlusskriterien, Literatursuche und die Zahl der eingeschlossenen Primärstudien zumindest in den drei komplementärmedizinischen Bereichen häufig und in hohem Ausmaß differierten.

Ein entscheidender Punkt in der hier präsentierten vergleichenden Analyse ist die Frage, was als *gleiche* Fragestellung bezeichnet werden kann. Hierzu dürften Forscher, die selbst systematische Reviews durchführen und der durchschnittliche Leser solcher Arbeiten völlig unterschiedliche Ansichten haben. Für den erfahrenen Reviewer ist klar, dass die Fragestellung letztendlich durch die Einschlusskriterien definiert wird und dass unterschiedliche Einschlusskriterien zu divergierenden Ergebnissen führen können. Der durchschnittliche Leser aber wird sich daran orientieren, was explizit als Ziel oder Fragestellung in Abstract oder Einleitung genannt ist. Von verschiedenen Reviews zu einer Frage wie „Sind Hypericumextrakte wirksamer als Placebo in der Behandlung der Depression?“ wird er sich – wenn er davon ausgeht, dass systematische Reviews zuverlässig sind – gleiche Antworten erwarten. Er wird kaum wissen, dass nur durch die Nichtberücksichtigung nicht in Englisch publizierter Studien drei Viertel der vorhandenen randomisierten Studien ignoriert werden. Aus der Sicht des Autors ist es angebracht, für eine vergleichende Analyse die Perspektive der Mehrheit der Leser einzunehmen, die die Ergebnisse eines Reviews perzipieren und interpretieren.

Ein Schwachpunkt der vorgelegten Analyse ist, dass aufgrund begrenzter Ressourcen die Analyse nur vom Autor allein durchgeführt wurde. Das ist bei einer qualitativen Auswertung besonders problematisch, da persönliche Sichtweisen hier einen großen Einfluss haben können.

In der Literatur gibt es eine Reihe von Veröffentlichungen, in denen Beispiele von systematischen Übersichtsarbeiten zum gleichen Thema mit divergierenden Ergebnissen diskutiert wurden (66;74;190;204;231). Keine dieser Untersuchungen kann aber Aussagen darüber machen, wie häufig Diskrepanzen in einer zufälligen oder repräsentativen und gleichzeitig ausreichend großen

Stichprobe von Fragestellungen sind. In der vorliegenden Analyse wurde alle 17 Fragestellungen zu Akupunktur, Phytotherapie und Homöopathie, bei denen mehr als zwei Reviews vorlagen, analysiert. In diesen Bereichen sind relevante Nichtübereinstimmungen eher die Regel als die Ausnahme. Bei 10 von 17 Fragestellungen variierte die Anzahl der eingeschlossenen Studien bzw. das Studienset um mehr als 50%. Dass dies nicht zu größeren Diskrepanzen bei den Schlussfolgerungen führte, liegt vermutlich einerseits an der generellen Tendenz, definitive Aussagen aufgrund von methodischen Schwächen der Primärstudien bzw. von zu geringen Studienzahlen zu vermeiden sowie daran, dass sich möglicherweise ein- und ausgeschlossene Studien bzgl. der Ergebnisse wenig unterschieden.

Es gibt überzeugende Evidenz, dass bei randomisierten Studien eine bessere methodische Qualität mit weniger positiven Ergebnissen assoziiert ist (287;331;406) (siehe auch Kapitel 3). Bzgl. systematischer Reviews sind die wenigen verfügbaren Ergebnisse über einen Zusammenhang zwischen Qualität und Schlussfolgerung widersprüchlich. Während Jadad und McQuay (204) fanden, dass weniger hochwertige Reviews häufiger positive Ergebnisse hatten, war bei Katerndahl und Lawler (231) das Gegenteil der Fall und Jadad et al. (205) fanden keinen Zusammenhang. In der vorliegenden Analyse scheint die Qualität nicht ein vorrangiger Grund für die Diskrepanzen zu sein. Z.B. gibt es beim Thema Hypericum bei Depression gute Argumente, entweder alle Studien (wie in (285)), nur Studien mit einer Diagnosestellung nach hochwertigen Kriterien (235) oder nur Studien mit mindestens 6 Wochen Behandlungsdauer (509) einzuschliessen. Solche fundamentalen Einschränkungen, die dazu führen, dass die jeweilige Übersicht sich von einer älteren deutlich unterscheidet, sollten jedoch bereits in der Einleitung einer Publikation explizit dargelegt werden.

Möglicherweise sind bei komplementärmedizinischen Interventionen aufgrund der häufig wenig spezifischen Fragestellungen Diskrepanzen häufiger wie in der konventionellen Medizin. Es erscheint jedoch sinnvoll, auch hier systematische vergleichende Analysen anzustellen.

Die vorgelegten Ergebnisse sollten nicht als Aufforderung missverstanden werden, wieder primär nichtsystematische, narrative Reviews durchzuführen. Ohne explizite, transparente und rigorose Methoden werden Übersichtsarbeiten sicher nicht besser. Es wäre aber irrig zu glauben, dass bei der Vielzahl von Entscheidungen, die bei der Erstellung eines Protokolls und während der Durchführung eines systematischen Reviews getroffen werden müssen, in jedem Fall zuverlässige und reproduzierbare Ergebnisse zu erwarten sind.

5 Schlussbetrachtung

In der vorliegenden Monographie wurden zwei umfangreiche eigene systematische Übersichtsarbeiten klinischer Studien (Kapitel 2), drei Untersuchungen zur methodischen Qualität und deren Auswirkungen auf die Studienergebnisse (Kapitel 3) sowie eine Analyse der vorliegenden systematischen Übersichten zu Akupunktur, Phytotherapie und Homöopathie (Kapitel 4) präsentiert.

An den beiden eigenen Übersichtsarbeiten werden die Stärken von systematischen Übersichtsarbeiten deutlich (100). Durch die umfangreiche Literatursuche wird gewährleistet, dass die potentiell relevanten Studien – zumindest wenn sie publiziert sind - weitgehend vollständig identifiziert werden. Die expliziten Ein- und Ausschlusskriterien und deren systematische Prüfung machen den Selektionsprozess transparent und gewährleisten eine unverzerrte Auswahl. Durch die standardisierte Extraktion und Beurteilung werden die eingeschlossenen Primärstudien unabhängig von ihren Ergebnissen ausgewertet und beurteilt. In tabellarischer Form sind Details jeder einzelnen Studie für den Leser zu überprüfen, ohne dass im Text endlose Beschreibungen der Einzelstudien erfolgen müssen. Die Berechnung und graphische Darstellung von Effektmaßen ermöglicht schließlich eine schnelle und intuitive Erfassung der vorliegenden Studienergebnisse. In der Gesamtschau liegt eine transparente und nachprüfbare Zusammenfassung der vorliegenden Untersuchungen zu den bearbeiteten Fragestellungen vor.

Jedoch treten auch die Probleme von systematischen Übersichtsarbeiten klar hervor (99;132). Ein Hauptproblem ist die Heterogenität der Primärstudien bzgl. Patienten (genaue Diagnosen, aus Praxen oder spezialisierten Zentren, Chronizität, Vorerfahrungen mit Therapien etc.), Interventionen (Dauer, Dosis etc.), Zielkriterien (Art, Zeitpunkt) und methodischer Qualität. Durch äußerst strikte Einschlusskriterien ist es möglich, diese Heterogenität zu minimieren. Dies ist sinnvoll, wenn die Fragestellung bis in Details sehr genau spezifizierbar ist. In diesen Fällen ist auch eine quantitative Meta-Analyse mit hoher Wahrscheinlichkeit sinnvoll und zielführend. Ein Beispiel für eine solche Fragestellung wäre „Gibt es Evidenz aus randomisierten, doppelblinden Studien, dass eine prophylaktische Behandlung mit Metoprolol mit Dosierungen zwischen 100 und 200 mg pro Tag bei Patienten mit Migräne ohne Aura die Attackenfrequenz stärker vermindert als Placebo?“ Für Ärzte und Patienten sind derartige Fragestellungen sinnvoll, wenn Sie konkrete Handlungsanweisungen für eine definierte Situation suchen. Die entsprechenden Übersichtsarbeiten enthalten meist nur sehr wenige Primärstudien. Systematische Übersichtsarbeiten mit weiteren Fragestellungen (z.B. „Gibt es Evidenz aus randomisierten Studien für die Wirksamkeit einer Kopfschmerzprophylaxe mit Metoprolol?“) – wie z.B. den Anwendungsbeispielen in Kapitel 2 – informieren den Leser in einem weiteren und allgemeineren Sinne. Unweigerlich werden dann jedoch heterogene Studien eingeschlossen. Der Charakter des

Reviews wird dementsprechend mehr deskriptiv im Sinne einer Zusammenfassung des state-of-the-art als hypothesenprüfend. Dennoch werden bei engen und weiten Fragestellungen weitgehend die gleichen Methoden verwendet. Aus der Sicht des Autors dieser Monographie ist es beispielsweise absolut sinnvoll, auch zu den äußerst heterogenen Studien zu Hypericumextrakten bei Depression eine quantitative Meta-Analyse durchzuführen. Der entscheidende Punkt ist hier die Interpretation des berechneten globalen Effektmaßes. Auf der Suche nach einfachen und griffigen Aussagen wird typischerweise dieses Maß als „das“ Ergebnis einer Übersicht angesehen. In einer Sekundärpublikation zu der früheren Version des in Kapitel 2 aktualisierten Hypericumreviews in der Zeitschrift *Evidence-Based Mental Health* (192) wurde z.B. sogar eine „number needed to treat“ berechnet, d.h. aus der gepoolten Responderratio wurde abgeleitet, wieviele Patienten mit Hypericum behandelt werden müssen, damit (im Vergleich zu Placebo) ein Patient profitiert. Dieses für Ärzte und Patienten intuitiv zu verstehende Maß wird von Vertretern der evidenzbasierten Medizin bevorzugt zur Ergebnisdarstellung bei prophylaktischen oder therapeutischen Interventionen verwendet (255). Angesichts der Heterogenität der Hypericum-Studien in Bezug auf Patienten, Extrakte und Studienqualität ist eine solche vereinfachende Darstellung potentiell irreführend. Gepoolte Effektmaße können in Reviews mit weiteren Fragestellungen nicht mehr sein als grobe Indikatoren dafür, ob die Evidenz insgesamt positiv ist oder nicht, bzw. ob es sich eher um einen großen Effekt oder einen kleinen Effekt handelt. Auch wenn keine gepoolten Effektmaße berechnet werden, bleibt die Heterogenität ein Problem, wenn aus dem Gesamteindruck der Evidenz abgeleitet wird, dass die Ergebnisse für Einzelsituationen quasi universell gültig seien.

Ein zweites Hauptproblem in vielen systematischen Reviews ist die häufig unzureichende Berichterstattung in den Publikationen der Primärstudien. Bei den Anwendungsbeispielen trat dieses Problem insbesondere im Fall der Migräneprophylaxe mit Propranolol, Metoprolol und Flunarizin auf. Hier war die Berichterstattung der Ergebnisse so inkonsistent, dass allenfalls durch vote counts (Klassifikation der Ergebnisse in Kategorien) und die deskriptive Ergebniszusammenfassung in den Tabellen im Anhang ein Gesamtbild der Ergebnisse entsteht. Vote counts sind jedoch nicht nur äußerst grobe Simplifizierungen, sondern – wie auch das Beispiel der Übersichten zur Akupunktur bei LWS-Schmerzen in Kapitel 4 zeigt - von fraglicher Zuverlässigkeit ((176) S. 48-52). Die deskriptiven Zusammenfassungen der einzelnen Studienergebnisse im Tabellenanhang sind für den „normalen“ Leser aufgrund ihres Umfangs äußerst ermüdend und primär für den besonders Interessierten „zum Nachsehen“ gedacht. Die graphische Darstellung der Studienergebnisse mit den einzelnen Effektgrößen wäre eindeutig die Methode der Wahl für eine informative Darstellung. Wenn jedoch nur für weniger als die Hälfte der Studien, die den jeweiligen Parameter gemessen haben sollten, Daten vorliegen, muss angezweifelt werden, dass diese Darstellung einen repräsentativen Eindruck vermittelt. Der Gedanke liegt nahe, dass Autoren die besten und eindeutigsten Daten besonders ausführlich präsentieren (outcome reporting bias (99)).

Bei beiden Beispielübersichtsarbeiten stellt sich auch die Frage, ob möglicherweise ein publication bias, also eine Überschätzung der Wirksamkeit aufgrund nichtpublizierter negativer Ergebnisse erfolgte (242). Eindeutige Hinweise lagen zwar nicht vor, angesichts des finanziellen Interesses an den jeweiligen Behandlungen kann ein publication bias aber auch nicht mit Sicherheit ausgeschlossen werden.

Die empirisch-methodischen Untersuchungen zu Studienqualität und evtl. Zusammenhängen mit den Studienergebnissen in Kapitel 3 verdeutlichen einerseits, wie grob und eindimensional die verfügbaren Instrumente zur Beurteilung der Studienqualität sind, andererseits kann aber sogar mit diesen Instrumenten klar gezeigt werden, dass methodisch bessere Studien offensichtlich im Schnitt weniger optimistische Ergebnisse erbringen als methodisch schlechtere Studien. Dies gilt sowohl für die hier untersuchten komplementärmedizinischen Beispiele wie auch in der konventionellen Medizin (331;406). Eine Beurteilung der methodischen Qualität muß daher trotz aller Probleme für systematische Übersichtsarbeiten als obligat angesehen werden. Methodische Weiterentwicklungen sind wünschenswert, eine schnelle Entwicklung einer optimalen Qualitätsbeurteilungsmethode erscheint jedoch wenig realistisch.

Die Ergebnisse in Kapitel 4 zeigen, dass viele systematische Übersichtsarbeiten zu Akupunktur, Phytotherapie und Homöopathie – ähnlich wie in untersuchten Bereichen der konventionellen Medizin (69;201;203) - erhebliche methodische Schwächen haben. Es ist zu hoffen, dass die Weiterentwicklungen der vergangenen Jahre, die Empfehlungen zur Durchführung von systematischen Reviews (2) und die Publikation von Leitlinien zur Publikation von systematischen Reviews und Meta-Analysen im QUORUM-Statement (327) in der Zukunft zu einer Verbesserung führen.

Eine zentrale Rolle wird hier der bereits in Kapitel 4 kurz erwähnten Cochrane Collaboration (www.cochrane.de) zufallen. Dieses zu Beginn der neunziger Jahre gegründete und nach dem britischen Arzt und Epidemiologen Archie Cochrane (1900-1988) benannte weltweite Netzwerk von Ärzten, Forschern, Patienten und Entscheidungsträgern hat es sich zur Aufgabe gemacht, systematische Übersichtsarbeiten der vorhandenen klinischen Studien in allen Bereichen der Medizin zu erstellen, kontinuierlich zu aktualisieren und zugänglich zu machen. In der von der Cochrane Collaboration auf CD-ROM herausgegebenen, vierteljährlich aktualisierten *Cochrane Library* (Oxford, Update Software) sind derzeit (Ausgabe 4, 2001) 1235 nach definierten Vorgaben erstellte und begutachtete „Cochrane Reviews“ verfügbar, darüberhinaus die Protokolle von 956 in Bearbeitung befindlichen Reviews sowie Referenzen und Abstracts von über 3000 anderen systematischen Reviews und 370.000 klinischen Studien. Im Rahmen dieses Netzwerkes wird in zunehmendem Maße auch empirisch-methodische Forschung zu systematischen Übersichtsarbeiten betrieben. Ein spezielles Register zu methodischen Untersuchungen wurde

angelegt (derzeit 3551 Zitate) und 12 Arbeitsgruppen geformt, die sich mit spezifischen Problemen beschäftigen.

Durch die Aktivitäten der Cochrane Collaboration, die große Zahl von in Zeitschriften publizierten Arbeiten in den letzten Jahren und den Trend zu „Evidenz-basierter Medizin“ haben systematische Übersichtsarbeiten auch für gesundheitspolitische Entscheidungen große Bedeutung erlangt. In der Hierarchie der Evidenz zu prophylaktischen und therapeutischen Interventionen wird Meta-Analysen randomisierter Studien häufig das größte Gewicht eingeräumt. Als noch zuverlässiger werden jedoch extrem große randomisierte Studien – sogenannten megatrials – angesehen (23). Eine methodisch-empirische Untersuchung fand, dass zwischen den Ergebnissen von Meta-Analysen und später zur selben Fragestellung durchgeführten megatrials nur eine mäßige Übereinstimmung bestand (274). Jedoch fand eine neuere Untersuchung, dass auch zwischen megatrials zur gleichen Fragestellung nur eine mäßige Übereinstimmung bestand (141).

Aus der Sicht des Autors dieser Monographie sollte in Zukunft stärker zwischen hypothesenprüfenden Meta-Analysen und systematischen state-of-the-art Reviews unterschieden werden. Hypothesenprüfende Meta-Analysen sind durchaus als eigene Evidenzklasse anzusehen und können – wenn die Primärstudien es zulassen – klare Antworten auf genau spezifizierte Fragen erbringen. State-of-the-art Reviews können und sollten ebenfalls systematische Methoden nutzen; ihr Ziel muss aber primär eine transparente Zusammenfassung der vorliegenden Evidenz sein. Es stellt sich die Frage, ob es angemessen ist, aus derartigen Übersichten klare Schlussfolgerungen zu ziehen. Die Ergebnisse sind nämlich häufig – entsprechend den weiten Fragestellungen - vielschichtig und wenig klar. Der Ermessensspielraum der Reviewautoren ist bereits bei der Zusammenfassung der Daten erheblich. Bei Schlussfolgerungen ist die Nachvollziehbarkeit aber letztlich nur noch in eingeschränktem Maße wirklich gegeben.

Obwohl bisher noch keine systematischen Untersuchungen darüber vorliegen, wie häufig in systematischen Übersichtsarbeiten die Fragestellungen „eng“ oder „weit“ sind (eine solche Untersuchung zu Cochrane Reviews ist derzeit im Gange (22)), ist der Anteil weitgefasster systematischer Übersichtsarbeiten sicher groß. Hier erschiene eine größere Bescheidenheit in der Darstellung und Interpretation angemessen. Für die jeweiligen Reviewautoren ist dies jedoch nicht einfach, da sich eine „klare“ und „griffige“ Schlussfolgerung besser publizieren lässt als eine deskriptive Zusammenstellung der Evidenz, die auf solche Schlussfolgerungen bewusst verzichtet.

In den kommenden 10 Jahren wird die methodische Entwicklung systematischer Übersichtsarbeiten mit Sicherheit weiter voranschreiten. Bei der Planung und Publikation neuer Primärstudien wird vermehrt daran gedacht werden, dass Ergebnisse später in Meta-Analysen zusammengeführt werden können. Im Optimalfall wird dadurch möglich, in Meta-Analysen auf die Originaldaten zugreifen (70), in anderen Fällen ist zumindest davon auszugehen, dass die

Ergebnisse häufiger in verwertbarer Weise präsentiert werden. Die Einrichtung von Registern, in denen Studien bereits bei der Durchführung erfasst werden (410), könnten das Problem eines publication bias vermindern. Die Erfahrung im Umgang mit den einzelnen Methoden und die empirisch-methodischen Untersuchungen werden dazu führen, dass unangemessene und ineffiziente Vorgehensweise eliminiert werden. Daher ist davon auszugehen, dass die Qualität zukünftiger Übersichtsarbeiten im Mittel deutlich besser sein wird als die der heute verfügbaren Reviews. Dennoch werden systematische Übersichtsarbeiten auch in Zukunft schwierige Vorhaben bleiben, deren Zuverlässigkeit von den Fähigkeiten, der Sorgfalt und den Erfahrungen der Reviewer abhängig sein wird.

6 Tabellenanhang

6.1 Tabellen zu den Studien zu Propranolol, Metoprolol und Flunarizin zur Migräneprophylaxe

Erklärung der Darstellungsweise und Abkürzungen

Autor	N	Diagnosen	%weiblich/ Alter/ Erkr.dauer	Design	Conceal./ Jadad score	Baseline/ Behandl./ Follow-up	Dosis	Kontrolle(n)
	rand./ anal.							

Spalte 1 Autor: Erstautor und Publikationsjahr (bei Jahren mit 19hundert nur die letzten 2 Ziffern)

Spalte 2: n rand. = Anzahl randomisierter Patienten, anal. = Zahl in die Auswertung einbezogener Patienten

Spalte 3 Diagnosen: In der Publikation genannte Diagnosen der Studienpatienten

Spalte 4: %weiblich = Anteil der weiblichen Patienten in Prozent, Alter = Alter in Jahren (Mittelwert oder Bereich je nach Verfügbarkeit), Erkr.dauer = Bestehensdauer der Migräne in Jahren

Spalte 5 Design: Studiendesign, entweder parallel (d.h. ein Patient erhält nur eine Studienbehandlung) oder cross-over (Patient erhält in verschiedenen Phasen unterschiedliche Studienbehandlungen)

Spalte 6: Conceal. = concealment, d.h. die Beurteilung der Methode, mit der sichergestellt wurde, dass der einschließende Arzt vor Aufnahme eines Patienten in die Studie nicht wußte, welche Behandlung dieser bekommen würde (adäq. = Methode adäquat); Jadadscore = Punktzahl bei den einzelnen Items des Jadadscores (Item 1 maximal 2 Punkte für Randomisation, Item 2 maximal 2 Punkte für Doppelblindbedingungen, Item 3 maximal 1 Punkt zur Beschreibung von Studienabbrüchen und Ausschlüssen)

Spalte 7: Baseline = Dauer der Vorbehandlungsbeobachtungsphase in Wochen (w) oder Monaten (m); Behandl. = Dauer der Behandlungsphase (bei cross-over Studien Angabe der Behandlungsphase pro Intervention x Anzahl der Interventionen, z.B. 2 x 8 w bedeutet, dass pro Behandlungsphase 8 Wochen behandelt wurde); Follow-up = Dauer der Beobachtung nach Abschluss der Behandlung

Spalte 8 Dosis: Tägliche Wirkstoffdosis

Spalte 9 Kontrolle(n): Art und ggfs. Dosis der Vergleichsintervention

Vote count	Responder	Andere Ergebnisse	Bemerkungen
---------------	-----------	-------------------	-------------

Spalte 10 Vote count: Grobkategorisierung der Ergebnisse in + = Prüfbehandlung signifikant besser als Kontrolle, (+) = Trend zugunsten der Prüfbehandlung, 0 = kein Unterschied, (-) = Trend zugunsten der Kontrollbehandlung, - = Kontrollbehandlung signifikant besser; Vote counts sind in jeder Tabelle jeweils nur für den in der Überschrift genannten Vergleich angegeben

Spalte 11 Responder: Therapieerfolge in Prüf- vs. Kontrollgruppe (in Prozent oder als Anzahl Therapieerfolge/Anzahl Behandelte)

Spalte 12 Andere Ergebnisse: Kurzzusammenfassung anderer Ergebnisse, nach Möglichkeit zu einem Kopfschmerzhäufigkeitsmaß

Spalte 13 Bemerkungen

k.A. = keine Angabe

Werte bei Häufigkeitsmaßen sind nach Möglichkeit Mittelwerte und Standardabweichungen. Der Zeitbezug bei Attackenhäufigkeit, Kopfschmerztagen oder Akutmedikationsverbrauch erfolgt, sofern nicht anders angegeben, jeweils auf die letzten 4 Therapiewochen, häufig im Vergleich zu den 4 Wochen vor Therapie

Tabelle A 1:
Randomisierte Studien Propranolol vs. Placebo zur Migräneprophylaxe

Autor	n rand./ anal.	Diagnosen	%weiblich/Alt er/ Erkr.dauer	Design	Conceal./ Jadad score	Baseline/ Behandl./ Follow-up	Dosis	Kontrolle(n)
Ahuja 89 (6)	26 k.A.	Migräne	46% 17-55 k.A.	Cross-over	k.A./ 1 – 1 – 0	- 2x8w -	120 mg	Placebo
Al-Qassab 93 (9)	45 30	27 mit, 17 ohne Aura	80% 18-65 1-49	Cross-over	k.A./ 1 – 1 – 1	4w 3x2m -	1) 160 mg 2) 80 mg	Placebo
Borgesen 74a, 74b, 76 (46;47)	45 30	15 mit, 15 ohne Aura	83% 18-59 1-50	Cross-over	k.A./ 1 – 1 – 0	4w 2x3m -	120 mg	Placebo
Dahlöf 87 (78)	28 28	20 mit, 8 ohne Aura	93% 18-60 ≥2	Cross-over	k.A./ 1 – 1 – 1	4w 2x1m 2x5m	120 mg	Placebo
Diamond 76 (82)	83 62	Mit und ohne Aura	81% 21-62 k.A.	Cross-over	k.A./ 1 – 1 – 0	- 2x4-8w -	80-160 mg	Placebo
Diener 96 (88)	214 174	58 mit, 156 ohne Aura	78% 39 19	Parallel	k.A./ 1 – 1 – 1	4w 16w -	80-120 mg	Placebo Cyclandelat 1200 mg
Forssman 76 (135)	40 32	Meist ohne Aura	87% 17-51 2-40	Cross-over	k.A./ 1 – 2 – 1	10w 2x12w -	240 mg	Placebo
Grotemeyer 87 (153)	30 24	Alle ohne Aura ("vasomotoris ch")	73% 36 k.A.	Cross-over	k.A./ 0 – 1 – 0	- 2x12w -	120 mg	Placebo
Holdorff 77 (187)	53 37	Meist ohne Aura	k.A. k.A. k.A.	Parallel	k.A./ 1 – 1 – 0	- 3m -	80-120 mg	Placebo
Johnson 86 (218)	29 17	10 mit, 19 ohne Aura	69% 22-80 4-50	Cross-over	adäq./ 1 – 2 – 1	1m 3x3m -	40 mg	Placebo Mefenaminsä ure
Kuritzky 87 (257)	38 37	7 mit, 24 ohne Aura	k.A. 17-53 14	Cross-over	k.A./ 1 – 0 – 0	- 2x8w -	160 mg	Placebo
Malvea 73 (306)	31 29	Alle ohne Aura	87% 25-57 k.A.	Cross-over	k.A./ 1 – 0 – 0	1m 2x6w -	k.A.	Placebo
Mikkelsen 86 (325)	39 31	10 mit, 21 ohne Aura	84% 15-65 k.A.	Cross-over	k.A./ 1 – 2 – 1	-/ 3x12w/-	120 mg	Placebo Tolfenaminsä ure
Nadelmann 86 (339)	64 41	35 ohne, 27 mit und ohne Aura	83%/ 18-60/k.A.	Cross-over	k.A./ 1 – 2 – 1	4w 6+2x12w -	80-320 mg	Placebo
Palferman 83 (357)	16 10	Migräne	80% 41 17	Cross-over	k.A./ 1 – 1 – 0	- 2x8 -	120 mg	Placebo

Vote count	Responder	Andere Ergebnisse	Bemerkungen
+	k.A.	In den 8 Wochen mit Propranolol $8,6 \pm 5,9$ Attacken, unter Placebo $14,5 \pm 13,1$. Index $20,7 \pm 16,8$ vs. $38,0 \pm 39,1$	Keine Angaben zu Drop-outs
0	k.A.	Mediane Attackenfrequenz 3,8 (180 mg), 3,8 (80 mg) und 3,2 (Placebo) Tage	15 Drop-outs. Propranololpräparat mit langsamer Freisetzung (1 Tagesdosis)
+	14/30 vs. 9/30 ($\geq 50\%$ Attackenreduktion)	Signifikant größere Attackenreduktion unter Propranolol	Hohe Drop-out-Rate (15/45), dreimal publiziert
+	k.A.	Signifikante größere Attackenreduktion und geringerer Medikationsbedarf unter Propranolol	Nur als erweiterter Abstract verfügbar. Langer Follow-up. Angeblich keine Drop-outs
+	34 Präferenz Propranolol, 17 Placebo	Unklare Angaben	Hohe Drop-out Rate (21), unklare Präsentation
(+)	33/78=42% vs. 17/55=31% ($\geq 50\%$ Attackenreduktion)	Signifikante Reduktion der Gesamtmigräredauer/Monat unter Propranolol ($34,6 \pm 58,8$ vs. $13,7 \pm 51,7$ h)	Sorgfältige Studie, akzeptable Drop-out Rate und intent-to-treat Analyse
+	11 ($\geq 50\%$ Attackenreduktion) unter Verum, k.A. zu Placebo	Signifikant geringerer Beschwerdeindex und Medikationsbedarf in der Verumphase	8 Drop-outs
(+)	4 Patienten hatten 50% weniger Attacken als unter Placebo	19 ± 16 vs. 22 ± 16 Kopfschmerzattacken innerhalb von 8 Wochen	Grenzwertige Drop-out Rate. Hypothese der Autoren: Pat. mit vasomotorischem Kopfschmerz sprechen schlechter an
0	8/13 vs. 7/14 ($\geq 50\%$ Indexreduktion)	Kopfschmerzindex zu Beginn in Placebogruppe höher, bei Behandlungsende keine Unterschiede	Inakzeptable Drop-out Rate
+	k.A.	Attackenzahl über 3 Monate $13,8 \pm 12,0$ vs. $20,1 \pm 18,0$ (Mefenamins. $12,9 \pm 10,8$); Gesamtattackendauer 115 ± 89 vs. 184 ± 157 (116 ± 137) Stunden	12 Drop-outs bzw. Ausschlüsse; sonst recht gute cross-over Studie
+	k.A.	Attackenhäufigkeit, Intensität und Kopfschmerzdauer signifikant geringer unter Propranolol	Nur als Abstract verfügbar, daher kaum interpretierbar
+	Präferenz: 16 Propranolol, 8 Placebo, 5 keine	Kopfschmerzindex unter Propranolol niedriger (keine Signifikanzberechnung), Medikationsverbrauch ähnlich	Berichterstattung entspricht nicht modernen Ansprüchen
+	k.A.	Attackenzahl und Attackendauer signifikant geringer unter Propranolol, kein signifikanter Unterschied bei Intensität	Grenzwertige Drop-out-Rate (8/39)
+	k.A.	Deutlich stärkerer Rückgang der Kopfschmerz- und Medikationsverbrauchsindices unter Propranolol	Hoher Drop-out, sehr detailliert berichtet, leider nur wenig transparente Indices
(+)	k.A.	21 vs. 24 von je 56 Tagen mit Kopfschmerz, Index 47 vs. 52	Migränepatienten nur Subgruppe in kleiner Studie mit unakzeptabel hohem Drop-out

Tabelle A1 (Fortsetzung)

Randomisierte Studien Propranolol vs. Placebo zur Migräneprophylaxe

Autor	Zahl rand./ anal.	Diagnosen	%weiblich/ Alter/ Erkr.dauer	Design	Conceal./ Jadad score	Baseline/ Behandl./ Follow-up	Dosis	Kontrolle(n)
Pradalier 89 & 90 (378-380)	55 41	8 mit, 61 ohne Aura, 5 beides	76% 37 17	Parallel	k.A./ 1 – 2 – 1	4w 12w -	160 mg	Placebo
Pita 73 (370)	7 8	4 mit, 5 ohne Aura	78% 23-39 1-27	Cross- over	k.A./ 1 – 1 – 1	- 2x2m -	160 mg	Placebo
Sargent 85 (393)	161 129	Mit und ohne Aura	79% 18-62 20	Parallel	k.A./ 1 – 1 – 0	2w 15w -	120 mg	Placebo Naproxen 1100 mg
Solomon 86 (418)	k.A. 15	Mit und ohne Aura	k.A. k.A. k.A.	Cross- over	k.A./ 0 – 1 – 0	k.A. 3x2m k.A.	120 mg	Placebo Verapamil 240 mg
Stensrud 76 (436)	20 19	3 mit, 17 ohne Aura	70% 15-60 k.A.	Cross- over	k.A./ 1 – 1 – 1	- 3x4w -	160 mg	Placebo
Stensrud 80 (438)	35 28	6 mit, 29 ohne Aura	69% 25-60 k.A.	Cross- over	k.A./ 1 – 1 – 0	- 2x6w -	160 mg	Placebo Atenolol 100 mg
Tfelt-Hansen (459)	84 96 80	Migraine	74% 40 21	Cross- over	k.A./ 1 – 2 – 0	4w 3x12w -	160 mg	Placebo Timolol 20 mg
Weber 72 (492)	25 19	6 mit, 13 ohne Aura	52% 19-61 k.A.	Cross- over	k.A./ 1 – 1 – 0	- 2x3m -	80 mg	Placebo
Wideroe 76 (506)	30 26	6 mit, 24 ohne Aura	87% 18-55 k.A.	Cross- over	k.A./ 1 – 1 – 0	- 2x3m -	160 mg	Placebo
Ziegler 87 & 97 (530;531)	54 30	Migraine	73%/ 22-57/k.A.	Cross- over	k.A./ 1 – 1 – 0	4-8w 3x8w -	80-240 mg	Placebo, Amitriptylin 50-150 mg

Die Studie von Diamond 87 (83) nicht in Tabelle aufgeführt, da sie aufgrund des ungewöhnlichen Designs nicht in diesem Format abbildbar ist.

Vote count	Responder	Andere Ergebnisse	Bemerkungen
+	k.a.	3,15 ± 0,77 vs. 6,41 ± 1,70 Kopfschmerzattacken/Monat	Hohe Drop-out-Rate (19/74 bzw. 14/55), keinerlei Placeboeffekt
+	k.A.	Signifikant weniger Attacken und weniger starke Attacken unter Verum	Sehr kleine Studie mit auffällig positiven Ergebnissen
(+)	2.80 ± 0.95 vs. 2.38 ± 0.96 (auf Skala 1=poor bis 4=excellent)	Rückgang der wöchentlich Kopfschmerzstage um 0,21 ± 1,86 vs. -0,25 ± 1,57	Mäßig berichtete Multicenterstudie mit nur geringen Effekten, z.T. widersprüchlich
+	k.A.	4.5 vs. 6.0 Attacken/Monat, Intensität 1.5 vs. 2.0 (jeweils p<0.05)	Nur als Abstract verfügbare, kleine cross-over Studie – nicht sicher interpretierbar
(+)	k.A.	5,0 vs. 6,1 Kopfschmerzstage, Index 7,5 vs. 12,3	Kein Tagebuch. Zusätzlicher Vergleich mit d-Propranolol (kein Effekt über Placebo hinaus)
(+)	k.A.	9,2 vs. 8,8 vs. 10,3 Kopfschmerzstage unter Propranolol, Atenolol und Placebo	Unzureichend beschriebene Ergebnisse
+	48/89 vs. 24/90 (>50% Attackenreduktion; 44/89 unter Timolol)	3,7 ± 3,4 vs. 4,8 ± 3,8 Attacken (3,4 ± 3,1 unter Timolol). Kein Unterschied bzgl. Attackendauer und -intensität	Gute und sorgfältig berichtete Studie
+	15/19 vs. 2/19 (> 50% Attacken- oder Intensitätsreduktion)	k.A.	Kleine cross-over Studie mit extrem positiven Ergebnissen
+	25/26 vs. 10/26 (> 50% Attackenreduktion)	Im Mittel 0,4 vs. 1,7 Attacken pro Monat	Nur Patienten eingeschlossen, die vorher bereits gut auf Propranolol angesprochen hatten
+	12/30 (> 50% Indexreduktion), k.A. zu Placebo	Indexreduktion auf 405 vs. 511 (429 unter Amitriptylin)	Hoher Drop-out (24/54), komplexes Design, unzureichende Ergebnispräsentation

Tabelle A 2:
Randomisierte Studien Propranolol vs. andere Medikamente zur Migräneprophylaxe: Calciumantagonisten

Autor	Zahl rand./ anal.	Diagnosen	%weiblich/ Alter/ Erkr.dauer	Design	Conceal./ Jadad score	Baseline/ Behandl./ Follow-up	Dosis	Kontrolle(n)
Albers 89 (10)	40 20	Migraine	89% 23-47 k.A.	Parallel	k.A./ 2 – 0 – 1	- 6m -	120-180 mg	Nifedipin 60-90 mg
Bonuso 98 (44)	50 42	Alle ohne Aura	68% 20-45 k.A.	Parallel	k.A./ 1 – 0 – 0	k.A. 2m 4m	80 mg	Flunarizin 10 mg
Bordini 97 (45)	52 45	Alle ohne Aura	91% 17-48 k.A.	Parallel	k.A./ 1 – 1 – 0	3w 17w 6w	60 mg	1. Flunarizin 10 mg 2. Flunarizin + Prop.
Diener 99 (86)	k.A. 783	Migräne	k.A. k.A. k.A.	Parallel	k.A./ 0 – 1 – 0	4w 16w -	160 mg	1. Flunarizin 5 mg 2. Flunarizin 10 mg
Formisano (134)	91 22 19	Mit und ohne Aura	55% 39 20	Parallel	k.A./ 1 – 0 – 1	4w 16w 4w	120 mg	Nimodipin 120 mg
Gawel 92 (143)	94 76	37 mit und 57 ohne Aura	90% 36 17	Parallel	k.A./ 1 – 1 – 0	1m 4m -	160 mg	Flunarizin 10 mg
Gerber 91 (146)	58 50	4 mit und 54 ohne Aura	81% 18-65 1-55	Parallel	k.A./ 1 – 1 – 0	2m 7m 1-2m	40-160 mg	Nifedipin 10-40 mg, Metoprolol 50-200 mg
Lücking 88 (297), Soyka 1990 (426)	434 336	Vorwiegend mit Aura	82% 42 k.A.	Parallel	k.A./ 1 – 1 – 0	- 4m -	120 mg	Flunarizin 10 mg
Lücking 88 (297), Soyka (424)	87 69	Vorwiegend mit Aura	74% 42 k.A.	Parallel	k.A./ 1 – 1 – 0	- 4m -	120 mg	Flunarizin 10 mg
Ludin 89 (298)	59 59	15 mit, 44 ohne Aura	71% 34 k.A.	Parallel	k.A./ 1 – 1 – 1	1m 4m -	120 mg	Flunarizin 10 mg
Scholz 87 (402)	109 83	9 mit, 74 ohne Aura	77% 40 19	Parallel	k.A./ 1 – 1 – 0	8w 24w 3m	160 mg	Metoprolol 200 mg Flunarizin 10 mg Nifedipin 40 mg Dihydroergot. 10 mg
Shimell 90 (409)	58 48	27 mit, 30 ohne Aura	70% 16-61 17	Parallel	k.A. 1 – 2 – 1	- 4m -	180 mg	Flunarizin 10 mg
Solomon 86 (418)	k.A. 15	Migräne mit und ohne Aura	k.A. k.A. k.A.	Cross- over	k.A./ 0 – 1 – 0	k.A. 3x2m k.A.	120 mg	Verapamil 240 mg Placebo

Vote count	Responder	Andere Ergebnisse	Bemerkungen
(+)	12/20 vs. 6/20	Größere Attackenreduzierung unter Propranolol	Hohe Drop-out Raten: 7 (Propranolol) und 13 (Nifedipin)
0	15/19 vs. 18/23 ($\geq 50\%$ Attackenreduktion)	4 Monate nach Absetzen 9/19 vs. 9/23 mit $\geq 50\%$ Attackenreduktion	Studie zum frontotemporalen Nitroglyc.-Test – nur Responderang. bzgl. Wirksamkeit
0 (-)	Patientenurteil gut – exzellent in allen 3 Gruppen 14/15	Rückgang der Attackenzahl pro 20 Tage von 2,8 auf 1,3 vs. 2,6 auf 1,2 (Flunarizin) vs. 2,9 auf 1,1 (Flunarizin + Propranolol)	Niedrige Propranololdosis, geringe Fallzahl
0 0	48% (> 50% Attackenred.) vs. 53% bzw. 46% (5 bzw. 10mg Flunarizin)	Attackenrückgang von 2,8 auf 1,7 vs. 2,9 auf 1,6 vs. 2,9 auf 1,8	Große Studie, nur als Abstract verfügbar – nicht beurteilbar
0	k.A.	Attackenhäufigkeit am Ende der Studie $2,5 \pm 1,9$ vs. $3,0 \pm 1,9$	Kleine Studie mit Messung von endokrinologischen Effekten
(-)	51% vs. 67% (Patientenurteil)	Reduktion von Attackenhäufigkeit (mehr unter Flunarizin nach 1 und 4 Monaten), Kopfschmerztage und Analgetikabedarf in beiden Gruppen	Akzeptable Drop-out Rate, keine intent-to-treat Analyse
(+)	6/19 Propranolol, 12/22 Metoprolol, 1/17 Nifedipin (Zeitreihenanalyse)	Rückgang von Migränetagen, Dauer und Intensität am größten unter Metoprolol, geringer unter Propranolol, Nifedipin ohne Wirkung	Sorgfältige Studie mit detaillierter Responderanalyse, aber ohne Varianzen bei deskriptiven Daten; Drop-out nicht völlig klar
0	105/170 vs. 104/166 (Globalurteil Untersucher)	Reduktion von Attackenhäufigkeit, Schmerzintensität und Analgetikagebrauch sehr ähnlich in beiden Gruppen	Hohe Ausschlussrate, keine intent-to-treat Analyse
0	24/34 vs. 24/35 (Globalurteil Untersucher)	Reduktion von Attackenhäufigkeit, Schmerzintensität und Analgetikagebrauch sehr ähnlich in beiden Gruppen	Hohe Ausschlussrate, keine intent-to-treat Analyse
0	13/27 vs. 16/32 (50% Attackenreduktion)	Mittelwerte für Attackenhäufigkeit, Medikationsverbrauch, Index und Dauer gehen unter Propranolol zurück, unter Flunarizin dagegen nicht	Exzellent berichtete Studie; bei Flun. entweder gute Response oder Verschlechterung, Propr. weniger Responder aber im Mittel besser
0 0 0	33% (Einzelfallanalyse) 60% 17% 31% 8%	30% (% Verminderung Kopfschmerztage) 54% 11% 37% 5%	Komplexe Studie, verwirrende Beschreibung von Drop-outs, schwer interpretierbar
0	94% vs. 70% guter oder sehr guter Effekt (global Patient)	Rückgang der Attacken von 5,7 auf 1,17 pro Monat vs. 4,6 auf 1,35; geringer Intensitätsrückgang in beiden Gruppen	Kein Kopfschmerztagebuch, unbefriedigende Ergebnispräsentation, ansonsten akzeptabel
0	k.A.	4,5 vs. 5,0 Attacken/Monat, Intensität 1,5 vs. 1,4, Dauer pro Attacke 4,3 vs. 3,3 h	Nur als Abstract verfügbare, kleine cross-over Studie – nicht sicher interpretierbar

Tabelle A 3:
Randomisierte Studien Propranolol vs. andere Medikamente zur Migräneprophylaxe: andere Betablocker

Autor	Zahl rand./ anal.	Diagnosen	%weiblich/ Alter/ Erkr.dauer	Design	Conceal./ Jadad score	Baseline/ Behandl./ Follow-up	Dosis	Kontrolle(n)
Gerber 91 (146)	58 50	4 mit, 54 ohne Aura	81% 18-65 1-55	Parallel	k.A./ 1 – 1 – 0	2m 7m 1-2m	40-160 mg	1. Metoprolol 50-200 mg, 2. Nifedipin 10- 40 mg
Hedman 86 (178)	12 k.A.	Mit Aura	67% 30-49 k.A.	Cross- over	k.A./ 1 – 1 – 0	2w 2x1m -	80 mg	Metoprolol 100 mg
Kangasniemi 84 (227)	36 33	6 mit, 30 ohne Aura	89% 18-51 16	Cross- over	k.A./ 1 – 2 – 1	4w 2x8w -	160 mg	Metoprolol 200 mg
Olerud 86 (350)	28 27	2 mit, 26 ohne Aura	79% 17-61 2-45	Parallel	k.A./ 1 – 1 – 1	4-17w 24w -	80-160 mg	Nadolol 40-160 mg
Olsson 84 (353)	56 53	22 mit, 34 ohne Aura	73% 19-59 5-43	Cross- over	k.A./ 1 – 2 – 1	4w 2x8w -	80 mg	Metoprolol 100 mg
Ryan 84 (392)	48 45	Migräne	73% 21-60 k.A.	Parallel	k.A./ 1 – 1 – 1	4w 12w -	120 mg	1. Nadolol 80 mg 2. Nadolol 160 mg
Scholz 87 (402)	109 83	9 mit, 74 ohne Aura	77% 40 19	Parallel	k.A./ 1 – 1 – 0	8w 24w 3m	160 mg	Metoprolol 200 mg Flunarizin 10 mg Nifedipin 40 mg Dihydroergot. 10 mg
Stensrud 80 (437;438)	35 28	6 mit, 29 ohne Aura	69% 25-60 k.A.	Cross- over	k.A./ 1 – 1 – 0	- 2x6w -	160 mg	Atenolol 100 mg Placebo
Sudilovsky 87 (445)	140 98	Mit und ohne Aura	76% 39 21	Parallel	k.A./ 1 – 2 – 0	4-8w 12w -	160 mg	1. Nadolol 80 mg 2. Nadolol 160 mg
Tfelt-Hansen 84 (459)	96 80	Migraine	74% 40 21	Cross- over	k.A./ 1 – 2 – 0	4w 3x12w -	160 mg	Timolol 20 mg Placebo

Vote count	Responder	Andere Ergebnisse	Bemerkungen
(-)	6/19 Propranolol, 12/22 Metoprolol, 1/17 Nifedipin (Zeitreihenanalyse)	Rückgang von Migränetagen, Dauer und Intensität am größten unter Metoprolol, geringer unter Propranolol, Nifedipin ohne Wirkung	Sorgfältige Studie mit detaillierter Responderanalyse, aber ohne Varianzen bei deskriptiven Daten; Drop-out nicht völlig klar
(?)	k.A.	Angriffshäufigkeit $2,4 \pm 0,3$ (Standardfehler, Propranolol) vs. $3,1 \pm 0,6$ (Metoprolol)	Kleine cross-over Studie mit Schwerpunkt auf Laborparametern
0	15/36 vs. 17/36 ($\geq 50\%$ Scorereduktion)	Angriffszahl $3,0 \pm 1,9$ vs. $3,0 \pm 1,8$; Score $5,4 \pm 3,5$ vs. $4,9 \pm 3,2$; Akutmedikation $4,7 \pm 3,8$ vs. $4,7 \pm 3,1$ Einnahmen	Gut beschriebene cross-over Studie mit geringer Drop-out Rate
0	8/14 vs. 5/13 ($\geq 50\%$ Angriffshäufigkeit)	In beiden Gruppen deutlicher Rückgang bei Angriffshäufigkeit und Analgetikabedarf, nicht bei Intensität und Dauer der einzelnen Angriffe	Kleine, aber sorgfältige Studie; jedoch deutliche Baselineunterschiede bei Angriffshäufigkeit und Analgetikabedarf
0	16/53 vs. 21/56 ($\geq 50\%$ Angriffshäufigkeit)	Angriffshäufigkeit von Baseline (5,4) auf 4,2 in beiden Gruppen, Intensitätsscore von 12,4 auf 8,7 vs. 9,7	Gut präsentierte Studie. Auffällige Exazerbation während Placebowashout zw. den Phasen
1. (-) 2. 0	k.A.	Angriffshäufigkeit um 2,88 vs. 3,39 (Nadolol 80 mg) bzw. 2,63 (Nadolol 160 mg)	Unbefriedigend berichtete Studie; Vorläuferstudie oder Untergruppe von Sudilovsky 1980?
(-)	33% (Einzelfallanalyse) 60% 17% 31% 8%	30% (% Verminderung Kopfschmerztag) 54% 11% 37% 5%	Komplexe Studie, verwirrende Beschreibung von Drop-outs, schwer interpretierbar
0	k.A.	9,2 vs. 8,8 vs. 10,3 Kopfschmerztag unter Propranolol, Atenolol und Placebo	Unzureichend beschriebene Ergebnisse
0 -	$\geq 50\%$ Angriffshäufigkeit: 5/44 vs. 13/49 vs. 17/47	Stärkerer Intensitätsrückgang und geringerer Medikationsverbrauch unter Nadolol 160 mg; Nadolol 80 mg und Propranolol 160 mg ähnlich	Sehr hohe Ausschlussrate
0	48/89 vs. 44/89 ($>50\%$ Angriffshäufigkeit; 24/90 unter Placebo)	$3,7 \pm 3,4$ vs. $3,4 \pm 3,1$ Angriffe ($4,8 \pm 3,8$ unter Placebo). Kein Unterschied bzgl. Angriffsdauer und -intensität	Gute und sorgfältig berichtete Studie

Tabelle A 4:
Randomisierte Studien Propranolol vs. andere Medikamente zur Migräneprophylaxe: andere Substanzen

Autor	Zahl rand./ anal.	Diagnosen	%weiblich/ Alter/ Erkr.dauer	Design	Conceal./ Jadad score	Baseline/ Behandl./ Follow-up	Dosis	Kontrolle(n)
Andersson 81 (18)	49 37	31 mit, 18 ohne Aura	69% 22-68 2-40	Cross- over	k.A./ 1 – 1 – 1	1m 2x3m -	160 mg	Femoxetin 400 mg
Baldrati 83 (26)	18 12	Ohne Aura	89% 18-49 3-38	Cross- over	k.A./ 1 – 1 – 1	1m 2x3m -	1,8 mg/kg	ASA 13,5mg/kg
Behan 80 (35)	56 36	12 mit, 44 ohne Aura	66% 18-56 1-33	Cross- over	k.A./ 0 – 1 – 0	- 2x3m -	120 mg	Methysergid 3 mg
Diener 96 (88)	214 174	58 mit, 156 ohne Aura	78% 39 19	Parallel	k.A./ 1 – 1 – 1	4w 16w -	80-120 mg	Cyclandelat 1200 mg Placebo
Gerber 95 (147) & Schellenberg 97 (395)	84/ 62	19 mit, 42 ohne Aura	90% 41 20	Parallel	k.A./ 1 – 1 – 0	8w 16w 12m	120-160 mg	Cyclandelat 1200- 1600 mg
Johnson 86 (218)	29 17	10 mit, 19 ohne Aura	69% 22-80 4-50	Cross- over	adäq./ 1 – 2 – 1	1m 3x3m -	40 mg	Mefenaminsäure Placebo
Kangasniemi 83 (228)	29 24	4 mit, 25 ohne Aura	86% 24-47 4-40	Cross- over	k.A./ 1 – 1 – 1	1m 2x3m 1m	160 mg	Femoxetin 400 mg
Kaniecki 97 (229)	37 32	Alle ohne Aura	81% k.A. k.A.	Cross- over	k.A./ 1 – 0 – 1	8w 2x12w -	60-240 mg	Valproinsäure 1000- 2000 mg
Kass 80 (230)	23/21	6 mit, 17 ohne Aura	70% 22-62 k.A.	Cross- over	k.A./ 1 – 1 – 0	4w 2x16w -	160 mg	Clonidin 100 µg
Kjaesgard- Rasmussen 94 (237)	76 56	14 mit, 62 ohne Aura	79% 19-65 1-40	Cross- over	k.A./ 1 – 2 – 1	4w 2x12w -	120 mg	Tolfenaminsäure 300 mg
Klapper 94 (238)	24/12	Migräne	k.A. k.A. k.A.	Cross- over	k.A./ 1 – 0 – 1	- 2x2m -	80-240 mg	Divalproat 750-1500 mg
Maissen 91 (304)	39 32	8 mit, 31 ohne Aura	67% 39 k.A.	Parallel	k.A./ 0 – 2 – 1	1m 4m 3m	120 mg	Hydroxytryptophan 300 mg
Mathew 81 (310)	340 273	Migräne (ohne andere Kopfschmerzen)	90% 35 k.A.	Parallel	k.A./ 1 – 0 – 1	1m 6m -	120-160 mg	1. keine Prophylaxe 2. Amitriptylin 50- 75mg (A) 3. Biofeedback (B) 4. Propranolol + B 5. A + B 6. Propranolol + A 7. Propranolol+ A + B

Vote count	Responder	Andere Ergebnisse	Bemerkungen
(+)	11/28 vs. 4/28	Attackenfrequenz $3,4 \pm 0,3$ (Standardfehler) vs. $4,1 \pm 0,3$, Index $16,8 \pm 2,3$ vs. $20,2 \pm 2,3$	Nur bei 28 Patienten Baselinedaten; 12 Drop-outs
0	9/18 vs. 9/18	Mittlere Indexreduktion in beiden Gruppen 65%	Hohe Drop-out rate (6), kleine Fallzahl
(+)	19/56 vs. 13/36	Attackenfrequenz geringer unter Propranolol, keine Unterschiede bzgl. Intensität	Hohe Drop-out Rate (20); keine Erwähnung von Randomisation
0	33/78=42% vs. 30/81=37% ($\geq 50\%$ Attackenreduktion)	Ähnliche Reduktion der Gesamtmigränedauer/Monat unter Propranolol ($34,6 \pm 58,8$ vs. $36,8 \pm 73,7$ h)	Sorgfältige Studie, akzeptable Drop-out Rate und intent-to-treat Analyse
0	Zeitreihenanalyse Migränetage: 11/42 vs. 12/42	Ähnliche Besserungen in beiden Gruppen, z.T. mit Vorteilen und weniger Nebenwirkungen bei Cyclandelat	Hoher Drop-out (22), keine intent-to-treat Analyse. Keine Varianzen bei deskriptiven Daten. Follow-up nicht interpretierbar (Drop-out)
0	k.A.	Attackenzahl über 3 Monate $13,8 \pm 12,0$ vs. $12,9 \pm 10,8$; Gesamtattackendauer 115 ± 89 vs. 116 ± 137 Stunden	12 Drop-outs bzw. Ausschlüsse. Sonst recht gute Cross-over-Studie
(+)	Präferenz: 11 Propranolol, 7 Femoxetin, 6 keine	Rückgang der Attackenfrequenz von $7,1 \pm 1,1$ auf $5,0 \pm 0,7$ vs. von $7,3 \pm 1,1$ auf $6,8 \pm 1,1$	Angemessen berichtete Studie, Ergebnisse für einzelne cross-over-Phasen separat präsentiert
0	20/37 vs. 21/37 ($\geq 50\%$ Attackenreduktion)	k.A.	Pragmatische cross-over Studie mit Berichtsschwächen
(+)	13/23 vs. 8/23 ($>50\%$ Attackenreduktion)	13 ± 12 vs. 13 ± 10 KS-Tage. Weniger Akutmedikationsverbrauch und Arbeitsausfallzeit unter Propranolol	Kleine cross-over Studie mit ungewöhnlichen Parametern und unkonventioneller Ergebnispräsentation
0	k.A.	Reduktion der Migränetage vs. Baseline vor cross-over $-1,7 \pm 1,9$ vs. $-0,6 \pm 1,6$ Tage, bzgl. Migränestunden $-15,7$ (90,9) vs. $6,9$ (48,8)	Hohe Drop-out-Rate. Die ansonsten vertrauenswürdige Studie wird durch eine fehlerhafte Statistikdarstellung in Frage gestellt
?	k.A.	Valproinsäure signifikant wirksamer (Kopfschmerzhäufigkeit), aber 9 Abbrüche wg. UAW versus 3 unter Propranolol	Insuffizient berichtete und daher nicht beurteilbare Studie. Extrem hohe Drop-out-Rate
0	7/20 vs. 8/19 ($> 50\%$ Attackenreduktion)	Rückgang Attacken/Monat von $7,1 \pm 5,8$ auf $4,4 \pm 4,0$ vs. $9,4 \pm 5,7$ auf $7,3 \pm 7,4$	Grenzwertige Drop-out-Rate, auffällig hohe Attackenfrequenz
+	k.A.	Mittlerer Indexrückgang mit Propranolol: 62% ohne Prophylaxe: 20% unter Amitriptylin: 42% unter Biofeedback: 35% unter Propranolol + Amitriptylin: 64% unter Propranolol + Biofeedback: 74% unter Amitriptylin + Biofeedback: 48% Propranolol + Amitriptylin + Biofeedback: 73%	Ungewöhnliche 8-armige Studie. Hohe und zwischen den Gruppen z.T. deutlich unterschiedliche Drop-out-Raten – daher nur eingeschränkte Interpretierbarkeit! Nur Ergebnisse zu Kopfschmerzindizes

Tabelle A4 (Fortsetzung)

Randomisierte Studien Propranolol vs. andere Medikamente zur Migräneprophylaxe: andere Substanzen

Autor	Zahl rand./ anal.	Diagnosen	%weiblich/ Alter/ Erkr.dauer	Design	Conceal./ Jadad score	Baseline/ Behandl./ Follow-up	Dosis	Kontrolle(n)
Mathew 81 (445)	375 281	Kombinations- kopfschmerz (Migräne Spannungs- kopfschmerzen)	95% 40 + k.A.	Parallel	k.A./ 1 – 0 – 1	1m 6m -	120-160 mg	1. keine Prophylaxe 2. Amitriptylin 50- 75mg (A) 3. Biofeedback (B) 4. Propranolol + B 5. A + B 6. Propranolol + A 7. Propranolol+ A + B
Mikkelsen 86 (325)	39 31	10 mit, 21 ohne Aura	84% 15-65 k.A.	Cross- over	k.A./ 1 – 2 – 1	- 3x12w -	120 mg	Tolfenaminsäure Placebo
Nicolodi 97 (343)	256 k.A.	Migräne	76% 35 4	Parallel	k.A./ 1 – 1 – 0	1m 3m -	60-120 mg	Methysergid 15-30 mg
Sargent 85 (393)	161 129	Migräne mit und ohne Aura	79% 18-62 20	Parallel	k.A./ 1 – 1 – 0	2w 15w -	120 mg	Naproxen 1100 mg Placebo
Scholz 87 (402)	109 83	9 mit, 74 ohne Aura	77% 40 19	Parallel	k.A./ 1 – 1 – 0	8w 24w 3m	160 mg	Metoprolol 200 mg Flunarizin 10 mg Nifedipin 40 mg Dihydroergot. 10 mg
Ziegler 87 & 93 (530;531)	54 30	Migräne	73% 22-57 k.A.	Cross- over	k.A./ 1 – 1 – 0	4-8w 3x8w -	80-240 mg	Amitriptylin 50-150 mg Placebo

Vote count	Responder	Andere Ergebnisse	Bemerkungen
(-)	k.A.	Mittlerer Indexrückgang mit Propranolol: 52% ohne Prophylaxe: 18% unter Amitriptylin: 60% unter Biofeedback: 48% unter Propranolol + Amitriptylin: 69% unter Propranolol + Biofeedback: 62% unter Amitriptylin + Biofeedback: 66% Propranolol + Amitriptylin + Biofeedback: 76%	Ungewöhnliche 8-armige Studie. Hohe und zwischen den Gruppen z.T. deutlich unterschiedliche Drop-out-Raten – daher nur eingeschränkte Interpretierbarkeit! Nur Ergebnisse zu Kopfschmerzindices
(-)	k.A.	Kein Unterschied bei Attacken und Gesamtdauer, Tolfenaminsäure im Trend besser bei Intensität und Analgetikabedarf	Grenzwertige Drop-out-Rate (8/39)
0	k.A.	Attackenzahl pro Monat $3,1 \pm 1,4$ vs. $3,1 \pm 1,4$; Attackenintensität (100 mm VAS) 58 ± 17 vs. 56 ± 17	Große Studie, die jedoch nur sehr kurz in einer Publikation mitberichtet wird
0	$2,80 \pm 0,95$ vs. $2,86 \pm 1,05$ (auf Skala 1=poor bis 4=excellent)	Rückgang der wöchentlich Kopfschmerzstage um $0,21 \pm 1,86$ vs. $-0,48 \pm 2,02$	Mäßig berichtete Multicenterstudie mit nur geringen Effekten, z.T. widersprüchlich
(+)	nach 33% 60% 17% 31% 8%	Einzelfallanalyse: % Verminderung Kopfschmerzstage: 30% 54% 11% 37% 5%	Komplexe Studie, verwirrende Beschreibung von Drop-outs, schwer interpretierbar
0	12/30 vs. 10/30 (> 50% Indexreduktion)	Indexreduktion auf 405 vs. 429 (511 unter Placebo)	Hohe Drop-out-Rate (24/54), komplexes Design, unzureichende Ergebnispräsentation

Tabelle A 5:
Randomisierte Studien Metoprolol vs. Placebo zur Migräneprophylaxe

Autor		Zahl rand./ anal.	Diagnosen	%weiblich/ Alter/ Erkr.dauer	Design	Conceal./ Jadad score	Baseline/ Behandl./ Follow-up	Dosis	Kontrolle(n)
Andersson 83 (17)		75 71	17 mit, 54 ohne Aura	85% 17-57 19	Parallel	k.A./ 1 – 1 – 1	4w 8w -	200mg	Placebo
Kangasniemi 87 (226)		77 74	Alle mit Aura	80% 17-64 2-51	Cross- over	k.A./ 1 – 2 – 1	4w 2x8w -	200 mg	Placebo
Langohr 85 (262)		63 36	13 mit, 23 ohne Aura	67% 24-60 5-41	Cross- over	k.A./ 1 – 1 – 0	6w 2x4w 6w	100 mg	Placebo
Steiner 88 (435)		59 59	12 mit, 42 ohne Aura, 5 beide	76% 19-55 k.A.	Parallel	k.A./ 1 – 2 – 1	4w 8w -	100 mg	Placebo

Vote count	Responder	Andere Ergebnisse	Bemerkungen
+	10/34 vs. 4/37 (>50% Scorered.), 15/34 vs. 6/37 (Patientenurteil)	Migräneattackenhäufigkeit $3,1 \pm 1,5$ vs. $4,6 \pm 2,3$, Score $12,0 \pm 7,7$ vs. $16,5 \pm 8,0$, Medikationsverbrauch $10,8 \pm 12,0$ vs. $20,1 \pm 18,2$	Praktisch kein Ansprechen auf Placebo, hoher Akutmedikationsverbrauch; gut berichtete Studie
+	38/71 vs. 21/71 (Patientenurteil)	Attackenhäufigkeit pro 4 Wochen 1,8 vs. 2,5 ($p < 0,004$), Intensität 1,9 vs. 2,0, Analgetikaverbrauch 1,9 vs. 4,4	Sehr gute Ergebnisbeschreibung
+	44% vs. 27% ($\geq 50\%$ Scorereduktion)	Symptomrückgang stärker unter Metoprolol	Uninterpretierbar: Drop-out-Rate 27/63, 5 Gruppen, verwirrende Präsentation
(+)	k.A.	Attackenhäufigkeit $3,2 \pm 2,0$ vs. $3,7 \pm 1,6$ ($p = 0,13$), Medikationsbedarf $8,2 \pm 5,7$ vs. $9,4 \pm 4,8$	Gut berichtete Studie, nach Studienende Umstellung der Non-Responder von Placebo auf Verum oder Dosiserhöhung

Tabelle A 6:
Randomisierte Studien Metoprolol vs. andere Medikamente zur Migräneprophylaxe

Autor	Zahl rand./ anal.	Diagnosen	%weiblich/ Alter/ Erkr.dauer	Design	Conceal./ Jadad score	Baseline/ Behandl./ Follow-up	Dosis	Kontrolle(n)
Diener 01 (89)	270 230	50 mit, 220 ohne Aura	81% 40 14	Parallel	k.A./ 1 – 1 – 0	4w+4w 16w -	200 mg	ASA 300 mg
Gerber 91 (146), Diener 89 (90)	58 50	4 mit und 54 ohne Aura	81% 18-65 1-55	Parallel	k.A./ 1 – 1 – 0	2m 7m 1-2m	50-200 mg	1. Propanolol 60-160 mg, 2. Nifedipin 10- 40 mg
Grotemeyer 88 (155)	29 24	Migräne, bei 10 mit Aura	79% 39 16	Cross- over	k.A./ 1 – 2 – 1	8w 2x3m -	200 mg	Flunarizin 10 mg
Grotemeyer 90 (154)	28 21	Alle ohne Aura	82% 31 10	Cross- over	k.A./ 0 – 2 – 1	8w 3x12w -	200 mg	ASA 1500 mg
Hedman 88 & 86 (177;178)	12 k.A.	Mit Aura	67% 30-49 k.A.	Cross- over	k.A./ 1 – 1 – 0	2w 2x1m -	100 mg	Propranolol 80 mg
Kangasniemi 84 (227)	36 33	6 mit, 30 ohne Aura	89% 18-51 16	Cross- over	k.A./ 1 – 2 – 1	4w 2x8w -	200 mg	Propranolol 160 mg
Langohr 85 (262)	63 36	13 mit, 23 ohne Aura	67% 24-60 5-41	Cross- over	k.A./ 1 – 1 – 0	6w 2x4w 6w	100 mg	Clomipramin 100 mg
Louis 85 (295)	33 31	5 mit, 26 ohne Aura	81% 18-57 2-40	Cross- over	k.A./ 1 – 2 – 1	2w 2x8w -	100 mg	Clonidin 100 µg
Olsson 84 (353)	56 53	22 mit, 34 ohne Aura	73% 19-59 5-43	Cross- over	k.A./ 1 – 2 – 1	4w 2x8w -	100 mg	Propranolol 100 mg
Scholz 87 (402)	109 83	9 mit, 74 ohne Aura	77% 40 19	Parallel	k.A./ 1 – 1 – 0	8w 24w 3m	160 mg	Propranolol 160 mg Flunarizin 10 mg Nifedipin 40 mg Dihydroergot. 10 mg
Sörensen 91 (422)	149 127	39 mit, 110 ohne Aura	79% 20-64 1-50	Parallel	k.A./ 1 – 2 – 1	4w 20w -	200 mg	Flunarizin 10 mg
Vilming 85 (477)	36 30	10 mit, 25 ohne Aura	83% 19-58 2-42	Cross- over	k.A./ 1 – 1 – 1	4w 2x8w -	100 mg	Pizotifen 1,5 mg
Wörz 92 & 91 (523;524)	125 78	34 mit, 91 ohne Aura	78% 39 20	Cross- over	k.A./ 1 – 1 – 0	4w 2x12w -	100-200 mg	Bisoprolol 5-10 mg

Vote count	Responder	Andere Ergebnisse	Bemerkungen
+	45% vs. 30% Responder (> 50% Attackenreduktion)	53 vs. 42 Patienten mit Nebenwirkungen, Attackenreduktion pro 4 Wochen von $3,6 \pm 1,2$ auf $1,8 \pm 1,6$ vs. von $3,4 \pm 1,3$ auf $2,4 \pm 1,9$	Vermutlich sehr gute, aber mäßig präsentierte Studie mit hoher Drop-out-Rate
1. (+) 2. +	12/22 Metoprolol, 6/19 Propranolol, 1/17 Nifedipin (Zeitreihenanal.)	Rückgang von Migränetagen, Dauer und Intensität am größten unter Metoprolol, geringer unter Propranolol, Nifedipin ohne Wirkung	Sorgfältige Studie mit detaillierter Responderanalyse aber ohne Varianzen bei deskriptiven Daten; Drop-out nicht völlig klar
0	12/24 vs. 7/24 ($\geq 50\%$ Attackenreduktion)	Im Mittel 53% vs. 56% Attackenreduktion im Vergleich zur Baseline	Sorgfältige Studie, leider nur basale Ergebnisse zur Attackenreduktion präsentiert
+	14/21 vs. 3/21 ($\geq 50\%$ Attackenreduktion)	Im Mittel $50 \pm 18\%$ vs. $26 \pm 22\%$ Attackenreduktion im Vergleich zur Baseline; keine Unterschiede bzgl. Intensität	Sorgfältige Studie mit relevantem Drop-out; keine Angabe ob randomisiert
(?)	k.A.	Attackenhäufigkeit $3,1 \pm 0,6$ (Standardfehler) vs. $2,4 \pm 0,3$	Kleine Cross-over Studie mit Schwerpunkt auf Laborparametern
0	17/36 vs. 15/36 ($\geq 50\%$ Scorerreduktion)	Attackenzahl $3,0 \pm 1,8$ vs. $3,0 \pm 1,9$; Score $4,9 \pm 3,2$ vs. $5,4 \pm 3,5$; Akutmedikation $4,7 \pm 3,1$ vs. $4,7 \pm 3,8$	Gut beschriebene Cross-over Studie mit geringer Drop-out Rate
+	44% vs. 24% ($\geq 50\%$ Scorerreduktion; 27% unter Placebo)	Symptomrückgang am stärksten unter Metoprolol; mehr Nebenwirkungen unter Clomipramin	Uninterpretierbar: Drop-out-Rate 27/63, 5 Gruppen, verwirrende Präsentation
+	10/31 vs. 8/31 ($\geq 50\%$ Scorerreduktion)	Attackenzahl von 5,5 auf 3,8 vs. 5,2, Tablettenverbrauch von 7,1 auf 5,1 vs. 8,0, Index von 11,2 auf 8,4 vs. 10,2	Gut berichtete Studie
0	21/56 vs. 16/53 ($\geq 50\%$ Attackenreduktion)	Attackenreduktion von Baseline (5,4) auf 4,2 in beiden Gruppen, Intensitätsscore von 12,4 auf 9,7 vs. 8,7	Gut präsentierte Studie. Auffällige Exazerbation während Placebowashout zwischen den Phasen
(+) (+) (+) +	Einzelfallanalyse: 60% 33% 17% 31% 8%	% Verminderung Kopfschmerztag: 54% 30% 11% 37% 5%	Komplexe Studie, verwirrende Beschreibung von Drop-outs, schwer interpretierbar
0	60% vs. 60% (Patientenurteil „definitively better“)	35% vs. 44% Attackenrückgang, aber 1 vs. 10 Studienabbrüche wegen Nebenwirkungen	Gute Studie, Ansprechraten insgesamt niedrig
(-)	10/34 vs. 14/33 ($\geq 50\%$ Scorerreduktion)	Attackenrückgang (/4 Wochen) von 6,0 auf 4,9 vs. 4,0, Analgetikabedarf von 7,6 auf 4,8 vs. 4,0	Gut berichtete Studie
0	40/78 vs. 41/78 ($\geq 50\%$ Attackenreduktion)	Attackenzahl $1,99 \pm 1,51$ vs. $2,05 \pm 1,99$, Intensität 2,40 vs. 2,38, Medikationsverbrauch 1,65 vs. 1,65	Gut präsentierte Studie mit unakzeptabel hoher Drop-out-Rate

Tabelle A 7:
Randomisierte Studien Flunarizin vs. Placebo zur Migräneprophylaxe

Autor	Zahl rand./ anal.	Diagnosen	%weiblich/ Alter/ Erkr.dauer	Design	Conceal./ Jadad score	Baseline/ Behandl./ Follow-up	Dosis	Kontrolle(n)
Al-Deeb 90 (8)	50 42	8 mit, 34 ohne Aura	76% 13-65 12	Parallel	k.A./ 0 – 1 – 0	6m? 3m -	10 mg	Placebo
Baker 87 (25)	58 k.A.	Mit und ohne Aura	81% k.A. k.A.	Parallel	k.A./ 0 – 1 – 0	4w 12w -	10 mg	Placebo
Diamond 93 (80)	143 101	33 mit, 68 ohne Aura	74% 35 17	Parallel	k.A./ 1 – 2 – 0	4w 20w -	10 mg	Placebo
Frenken 1984 (137;138)	k.A. 35	14 mit, 21 ohne Aura	83% 20-51 1-35	Parallel	k.A./ 1 – 2 – 0	k.A. 12w -	10 mg	Placebo
Louis 1981 (294)	58 58	18 mit, 30 ohne Aura	50% 20-47 2-20	Parallel	k.A./ 1 – 2 – 0	- 3m -	20 mg	Placebo
Mendenopoulos 85 (323)	30 20	Alle mit Aura	80% 20-65 5	Parallel	k.A. 2 – 1 – 0	1m 4m -	10 mg	Placebo
Paterna 90 (362)	50 34	Mit und ohne Aura	64% 18-60 k.A.	Cross- over	k.A./ 1 – 1 – 0	1m 5x3m -	10 mg	1. Placebo 2. Verapamil 240 mg 3. Diltiazem 180 mg 4. Nomodipin 90 mg
Pini 1985 (367)	k.A. 29	Mit und ohne Aura	83% 40 >5	Parallel	k.A./ 0 – 1 – 0	2m 17w -	10 mg	Placebo
Sörensen 86 (422)	29 27	Alle ohne Aura	79% 19-63 2-30	Cross- over	k.A./ 1 – 1 – 1	4w 2x16w -	10 mg	Placebo
Thomas 91 (460)	29 15	Alle ohne Aura	87% 20-43 k.A.	Cross- over	k.A./ 0 – 2 – 1	k.A. 2x3mo -	10 mg	Placebo

Vote count	Responder	Andere Ergebnisse	Bemerkungen
0	k.A.	Rückgang der Attackenhäufigkeit/3 Monate von 14 ± 8 auf 10 ± 5 vs. 10 ± 5 auf 8 ± 6 ($p = 0.08$)	Keine Aussage zu Randomisation, unklar ob wirklich 6 Monate Baseline vorangingen; Studie aus Saudi-Arabien
+	signifikant besseres Globalurteil (Arzt und Patient)	Signifikant weniger Migräneattacken unter Flunarizin	Nur als Abstract verfügbar – nicht sicher beurteilbar
(+)	20/50 vs. 17/51 ($\geq 50\%$ Attacken- oder Intensitätsreduktion)	Attackenrückgang von 4,2 auf 1,9 vs. 4,4 auf 2,7, Rückgang Medikationsindex 14 vs. 3	Uninterpretierbar: von 143 Patienten 101 analysiert, aber nur von 50 am Ende Daten
+	14/17 vs. 10/18 (Patientenurteil)	Attackenhäufigkeit nach 3 Monaten 0,8 vs. 2,6	Keine Erwähnung von Dropouts
+	25/29 vs. 13/29 (Patientenurteil)	Deutlich ausgeprägtere Attackenreduktion unter Verum; kein Effekt auf Intensität und Dauer	Keine Baselinephase, keine Drop-outs. Auffällig positive Ergebnisse
+	5/9 vs. 1/11 ($\geq 50\%$ Attackenreduktion)	Rückgang von Attackenfrequenz, Intensität und Index deutlich ausgeprägter unter Flunarizin	Hohe Drop-out-Rate
+	30/34 (keine Definition von Response) 19/34 25/34 34/34 32/34	Kopfschmerzindex 88 ± 122 287 ± 163 222 ± 161 74 ± 69 109 ± 117	Ungewöhnliche Studie mit 5-fach Cross-over und 20 Monaten Dauer. Dafür mit 16 Patienten geringe Drop-out-Rate, dennoch für die Analyse problematisch
?	k.A.	Kopfschmerzindex von $0,54 \pm 0,26$ auf $0,31 \pm 0,38$ vs. $0,33 \pm 0,26$ auf $0,30 \pm 0,32$	Primär elektronystagmographische Ergebnisse – Wirksamkeitsparameter unzureichend beschrieben
+	k.A.	Attackenfrequenz 2,0 vs. 3,2 (Baseline 3,5), nichtsignifikant weniger Akutmedikation, Intensität gleich	Sorgfältige Studie mit sehr technischer und verwirrender Ergebnispräsentation
(+)	5/15 vs. 2/15 ($\geq 50\%$ Attackenreduktion)	Rückgang Häufigkeitsindex von (Median) 4,0 auf 3,0 vs. 3,3. Intensitätsreduktion nur unter Flunarizin	Hohe Drop-out-Rate

Tabelle A 8:
Randomisierte Studien Flunarizin vs. andere Medikamente zur Migräneprophylaxe: Betablocker

Autor	Zahl rand./ anal.	Diagnosen	%weiblich/ Alter/ Erkr.dauer	Design	Conceal./ Jadad score	Baseline/ Behandl./ Follow-up	Dosis	Kontrolle(n)
Bonuso 98 (44)	50 42	Alle ohne Aura	68% 20-45 k.A.	Parallel	k.A./ 1 – 0 – 0	k.A. 2m 4m	10 mg	Propranolol 80 mg
Bordini 97 (45)	52 45	Alle ohne Aura	91% 17-48 k.A.	Parallel	k.A./ 1 – 1 – 0	3w 17w 6w	10 mg	1. Propranolol 60 mg 2. Flunarizin + Prop.
Diener 99 (86)	783 k.A.	Migräne	k.A. k.A. k.A.	Parallel	k.A./ 0 – 1 – 0	4w 16w -	1. 5 mg 2. 10 mg	Propranolol 160 mg
Gawel 92 (143;144)	94 71	37 mit, 57 ohne Aura	90% 36 17	Parallel	k.A./ 1 – 1 – 0	1m 4m -	10 mg	Propranolol 160 mg
Grotemeyer 88 (155)	29 24	Migräne, bei 10 mit Aura	79% 39 16	Cross- over	k.A./ 1 – 2 – 1	8w 2x3m -	10 mg	Metoprolol 200 mg
Lücking 88 (297), Soyka (426)	434 336	Vorwiegend mit Aura	82% 42 k.A.	Parallel	k.A./ 1 – 1 – 0	- 4m -	10 mg	Propranolol 120 mg
Lücking 88 (297), Soyka (424)	87 69	Vorwiegend mit Aura	74% 42 k.A.	Parallel	k.A./ 1 – 1 – 0	- 4m -	10 mg	Propranolol 120 mg
Ludin 89 (298)	59 59	15 mit, 44 ohne Aura	71% 34 k.A.	Parallel	k.A./ 1 -1 -1	1m 4m -	10 mg	Propranolol 120 mg
Scholz 87 (402)	109 83	9 mit, 74 ohne Aura	77% 40 19	Parallel	k.A./ 1 – 1 – 0	8w 24w 3m	160 mg	Metoprolol 200 mg Propranolol 160 mg Nifedipin 40 mg Dihydroergot. 10 mg
Shimell 90 (409)	58 48	27 mit, 30 ohne Aura	70% 16-61 17	Parallel	k.A./ 1 – 2 – 1	- 4m -	10 mg	Propranolol 180 mg
Sörensen 91 (422)	149 127	39 mit, 110 ohne Aura	79% 20-64 1-50	Parallel	k.A./ 1 – 2 – 1	4w 20w -	10 mg	Metoprolol 200 mg

Vote count	Responder	Andere Ergebnisse	Bemerkungen
0	18/23 vs. 15/19 ($\geq 50\%$ Attackenreduktion)	4 Monate nach Absetzen 9/23 vs. 9/19 mit $\geq 50\%$ Attackenreduktion	Studie zum frontotemporalen Nitroglycerin-Test – nur Responderangaben bzgl. Wirksamkeit
0	Patientenurteil gut – exzellent in allen 3 Gruppen 14/15	Rückgang der Attackenzahl pro 20 Tage von 2,6 auf 1,2 vs. 2,8 auf 1,3 (Propranolol) vs. 2,9 auf 1,1 (Flunarizin + Propranolol)	Niedrige Propranololdosis, kleine Fallzahl
0	53% (10 mg) vs. 46% (5 mg) vs. 48% ($> 50\%$ Attackenreduktion)	Attackenrückgang von 2.9 auf 1.6 vs. 2.9 auf 1.8 vs. 2.8 auf 1.7	Große Studie, nur als Abstract verfügbar – nicht beurteilbar
(+)	67% vs. 51% (Patientenurteil)	Reduktion von Attackenhäufigkeit (mehr unter Flunarizin nach 1 und 4 Monaten), Kopfschmerztagen und Analgetikabedarf in beiden Gruppen	Akzeptable Drop-out Rate, keine intent-to-treat Analyse
0	7/24 vs. 12/24 ($\geq 50\%$ Attackenreduktion)	Im Mittel 56% vs. 53% Attackenreduktion im Vergleich zur Baseline	Sorgfältige Studie, leider nur basale Ergebnisse zur Attackenreduktion präsentiert
0	104/166 vs. 105/170 (Globalurteil Untersucher)	Reduktion von Attackenhäufigkeit, Schmerzintensität und Analgetikagebrauch sehr ähnlich in beiden Gruppen	Hohe Ausschlussrate, keine intent-to-treat Analyse
0	24/35 vs. 24/34 (Globalurteil Untersucher)	Reduktion von Attackenhäufigkeit, Schmerzintensität und Analgetikagebrauch sehr ähnlich in beiden Gruppen	Hohe Ausschlussrate, keine intent-to-treat Analyse
0	13/27 vs. 16/32 (50% Attackenreduktion), 19/27 vs. 18/32 (Globalurteil)	Mittelwerte für Attackenhäufigkeit, Medikationsverbrauch, Index und Dauer gehen unter Propranolol zurück, unter Flunarizin dagegen nicht	Gute Studie; ungewöhnlich: Unter Flunarizin entweder gute Response oder Verschlecht., bei Prop. weniger Responder aber im Mittel besser
(-)	Einzelfallanalyse: 17%	% Verminderung Kopfschmerztag: 11%	Komplexe Studie, verwirrende Beschreibung von Drop-outs, schwer interpretierbar
0	60%	54%	
0	33%	30%	
0	31%	37%	
0	8%	5%	
0	70% vs. 94% guter oder sehr guter Effekt (global Patient)	Rückgang der Attacken von 4,6 auf 1,3 pro Monat vs. 5,7 auf 1,2, geringer Intensitätsrückgang in beiden Gruppen	Kein Kopfschmerztagebuch, unbefriedigende Ergebnispräsentation, ansonsten akzeptabel
0	60% vs. 60% (Patientenurteil „definitively better“)	44% vs. 35% Attackenrückgang, aber 10 vs. 1 Studienabbruch wegen Nebenwirkungen	Gute Studie, Ansprechraten insgesamt niedrig

Tabelle A 9:
Randomisierte Studien Flunarizin vs. andere Medikamente zur Migräneprophylaxe: andere Calciumantagonisten und Pizotifen

Autor	Zahl rand./ anal.	Diagnosen	%weiblich/ Alter/ Erkr.dauer	Design	Conceal./ Jadad score	Baseline/ Behandl./ Follow-up	Dosis	Kontrolle(n)
Bussone 87 (54)	30 25	Migräne	83% 36 13	Parallel	k.A./ 1 – 1 – 1	4w 12w -	10 mg	Nimodipin 120 mg
Centonze 85 (63)	30 27	25 ohne Aura, 2 ophtalmologische Migräne	72% 21-56 k.A.	Parallel	k.A./ 1 – 0 – 1	1m 3m -	10 mg	Verapamil 240 mg
Lamsudin 93 (261)	83 78	13 mit, 65 ohne Aura	41% k.A. k.A.	Parallel	k.A./ 1 – 1 – 0	1m 3m -	10 mg	Nifedipin 20 mg
Lastilla 90 (264)	k.A. 38	4 mit, 34 ohne Aura	79% 34 k.A.	Parallel	k.A./ 1 – 0 – 0	1m 3m -	10 mg	Verapamil 240 mg
Nuti 96 (345)	50 42	10 mit, 40 ohne Aura	68% 35 14	Parallel	k.A./ 1 – 0 – 1	4w 8m >6m	10 mg	Nimodipin 120 mg
Paterna 90 (362)	50 34	Mit und ohne Aura	64% 18-60 k.A.	Cross- over	k.A./ 1 – 1 – 0	1m 5x3m -	10 mg	1. Placebo 2. Verapamil 240 mg 3. Diltiazem 180 mg 4. Nimodipin 90 mg
Scholz 87 (402)	109 83	9 mit, 74 ohne Aura	77% 40 19	Parallel	k.A./ 1 – 1 – 0	8w 24w 3m	160 mg	Metoprolol 200 mg Propranolol 160 mg Nifedipin 40 mg Dihydroergot. 10 mg
Cerbo 85 (65)	30 27	Mit und ohne Aura	47% 23-54 >2	Parallel	k.A./ 1 – 1 – 0	4w 2m -	15 mg	Pizotifen 1,5 mg
Cerbo 86 (64)	27 21	Migräne	k.A. k.A. k.A.	Cross- over	k.A./ 0 – 1 – 0	1m 2x2m -	15 mg	Pizotifen 1,5 mg
Gelmers (zitiert nach Amery 83 (13))	39 38?	Mit oder ohne Aura	k.A. k.A. k.A.	Parallel	k.A./ 1 – 1 – 0	k.A. 3m k.A.	10 mg	Pizotifen (Dosis k.A.)
Louis 1982 (296)	75 72	28 mit, 44 ohne Aura	56% 17-54 2-32	Parallel	k.A./ 1 – 1 – 0	- 4m -	10 mg	Pizotifen 2-3 mg
Rascol 86 (382)	35 32	3 mit, 32 ohne Aura	71% 28-59 k.A.	Parallel	k.A./ 2 – 2 – 1	- 4m -	10 mg	Pizotifen 2,2 mg
Wörz 83 (522)	44 30	Migräne	75% 21-51 k.A.	Parallel	k.A./ 0 – 2 – 0	- 3m -	10 mg	Pizotifen 2,2 mg

Vote count	Responder	Andere Ergebnisse	Bemerkungen
(-)	6/13 vs. 12/14 (> 50% Attackenred.)	Rückgang von Medikamentenverbrauch und Kopfschmerzindex in beiden Gruppen, mehr Nebenwirkungen unter Flunarizin	Sehr kleine Studie, größere Streuung der Werte unter Flunarizin (Ausreißer)
(+)	k.A.	Rückgang der Attackenzahl von $4,3 \pm 2,3$ auf $1,8 \pm 2,1$ vs. von $5,3 \pm 2,4$ auf $3,8 \pm 1,5$	Z.T. deutliche Unterschiede in den Baselinewerten (v.a. Medikamentenverbrauch)
0	k.A.	Rückgang des Index von 26 ± 25 auf 7 ± 14 vs. 26 ± 26 auf 9 ± 14	Nur Angaben zu Index, ungewöhnlich hoher Männeranteil
0	k.A.	Rückgang der Attackenzahl von $8,1 \pm 2,9$ auf $2,0 \pm 2,1$ vs. von $7,4 \pm 4,6$ auf $1,8 \pm 1,5$. Weniger Nebenwirkungen unter Verapamil	Unklare Drop-out-Rate, angeblich einfachblind, sehr hohe Ansprechraten, gute Ergebnispräsentation
+	k.A.	Attackenrückgang von $5,0 \pm 2,4$ auf $1,3 \pm 1,2$ vs. von $7,4 \pm 2,9$ auf $3,7 \pm 3,1$; Anhalten des Therapieeffektes 8,4 vs. 4,9 Monate	Gute, allerdings nicht doppelblinde Studie, genaue Länge des Follow-up unklar
(+)	30/34 (keine Definition von Response) 19/34 25/34 34/34 32/34	Kopfschmerzindex 88 ± 122 287 ± 163 222 ± 161 74 ± 69 109 ± 117	Ungewöhnliche Studie mit 5-fach Cross-over und 20 Monaten Dauer. Dafür mit 16 Patienten geringe Drop-out-Rate, dennoch für die Analyse problematisch
0	Einzelfallanalyse: 17% 60% 33% 31% 8%	% Verminderung Kopfschmerztag: 11% 54% 30% 37% 5%	Komplexe Studie, verwirrende Beschreibung von Drop-outs, schwer interpretierbar
0	k.A.	$3,6 \pm 3,6$ vs. $2,7 \pm 2,4$ Attacken/Monat, 12 ± 7 vs. 14 ± 14 Stunden mit starkem Kopfschmerz pro Monat	Keine Baselinewerte für Zielparameter, Ergebnisse lassen Ungleichheiten vermuten
(+)	95% vs. 84% (> 50% Attackenred.)	$2,3 \pm 1,8$ vs. $4,2 \pm 2,4$ Attacken, Index 12 ± 6 vs. 18 ± 4	Randomisation nicht erwähnt, Verblindung wenig adäquat, Drop-outs nicht völlig klar
(+)	2/18 vs. 10/20 Patienten ohne Ansprechen auf Therapie	Attackenfrequenz im Vergleich zu vor der Behandlung $-36 \pm 12\%$ vs. $2 \pm 16\%$	Unpublizierte Studie zitiert als persönliche Mitteilung in einer Übersichtsarbeit
0	30/38 vs. 28/34 ($\geq 50\%$ Attackenred.); 22/38 vs. 16/34 (Patientenurteil)	Attackenreduktion über alle 4 Monate 54% vs. 45%, im 4. Monat in beiden Gruppen je 26 Pat. attackenfrei. Intensitätsreduktion unter Flunarizin	Multicenterstudie bei Allgemeinmed. mit sehr hohen Responderraten bei Patienten mit relativ geringer Attackenintens. Keine Baselinephase
(+)	Arztbeurteilung sehr effektiv 9/21 vs. 7/14	Rückgang der Attackenfrequenz 65% vs. 45%	Gute Studie, mäßige Ergebnispräsentation
0	15/21 vs. 13/21 (Arzturteil)	Attackenrückgang von 7 auf 4 vs. 7 auf 1,5 Attacken, mehr Intensitätsrückgang unter Flunarizin	Hohe Drop-out-Rate, kein Tagebuch, keine Erwähnung von Randomisierung

Tabelle A 10:
Randomisierte Studien Flunarizin vs. andere Medikamente zur Migräneprophylaxe: andere Substanzen

Autor		Zahl rand./ anal.	Diagnosen	%weiblich/ Alter/ Erkr.dauer	Design	Conceal./ Jadad score	Baseline/ Behandl./ Follow-up	Dosis	Kontrolle(n)
Agnoli 91 (3)		104 83	Alle ohne Aura	72% k.A. k.A.	Parallel	k.A./ 1 – 1 – 0	1m 6m 3m	5 mg	Dihydroergokryptin 10 mg
Drillisch 80 (97)		30 27	Migräne	83% 37 15	Parallel	k.A./ 0 – 1 – 0	6w 12w -	9 mg	Cinnarizin 225 mg
Gerber 89 (145)		30 k.A.	Mit und ohne Aura	k.A. k.A. k.A.	Parallel	k.A./ 1 – 1 – 0	6+4+2w 3m 6w	10 mg	Etilefrinpilvalat 20 mg
Grotemeyer (157)	87	34 29	Alle ohne Aura	56% 33 k.A.	Parallel	k.A./ 1 – 1 – 1	8w 12w -	5 mg	Etilefrinpilvalat 20 mg
Langohr (263)	1988	31 25	5 mit, 20 ohne Aura	88% 20-64 1-47	Parallel	k.A./ 1 – 2 – 0	8w 4m -	10 mg	Dihydroergotamin 10 mg
Mitsikostas (326)	97	44 41	9 mit, 35 ohne Aura	70% 35 k.A.	Parallel	k.A./ 1 – 0 – 1	4w 8w -	10 mg	Valproinsäure bis 1 g
Nappi 87 (341)		40 35	Alle ohne Aura	k.A. k.A. k.A.	Parallel	k.A./ 1 – 0 – 0	2m 3m -	5 mg	Cyclandelat 800 mg
Scholz 87 (402)		109 83	9 mit, 74 ohne Aura	77% 40 19	Parallel	k.A./ 1 – 1 – 0	8w 24w 3m	160 mg	Metoprolol 200 mg Propranolol 160 mg Nifedipin 40 mg Dihydroergot. 10 mg
Steardo 86 (433)		104 k.A.	9 mit, 95 ohne Aura	59% 30 11	Parallel	k.A./ 1 – 2 – 0	1m 5m -	10 mg	4mg Methysergid

Vote count	Responder	Andere Ergebnisse	Bemerkungen
(-)	44% vs. 64% (Definition unklar)	Deutlicherer und anhaltenderer Rückgang unter Dihydroergokryptin	Nur als Abstract verfügbar – nur eingeschränkt interpretierbar
?	Nur Angaben für beide Gruppen zusammen	Attackenrückgang von 12 auf 7 vs. 9 auf 4, geringerer Medikamentenverbrauch bei 10/14 vs. 8/14	Kleine Studie mit wenig befriedigender Ergebnispräsentation, unklar ob randomisiert
(+)	26% vs. 15% (keine Definition)	k.A.	Nur als Abstract verfügbar – nicht interpretierbar
0	k.A.	42 ± 22% vs. 56 ± 23% Rückgang der Kopfschmerzhäufigkeit im Vergleich zur Baseline	Nur als Abstract verfügbar – kaum interpretierbar
0	k.A.	Rückgang der Migränetage von 5,4 auf 4,7 vs. 6,0 auf 3,8; ähnlicher Rückgang bzgl. Schmerzintensität und Akutmedikationsverbrauch	Keine klare Schlussfolgerung der Autoren
0	14/20 vs. 15/21 (≥ 50% Attackenreduktion)	Attackenreduktion von 4,4 auf 1,8 vs. 5,1 auf 1,6, Akutmedikationsbedarf von 4,9 auf 2,3 vs. 4,4 auf 1,7	Sorgfältige, gute berichtete, aber offene Studie; mehr Dysthymiesymptome unter Flunarizin
+	k.A.	Rückgang von 11 auf 5 Tage mit Migräne vs. 12 auf 7, bei Medikation von 15 auf 6 vs. 15 auf 9	Angeblich doppelblind, jedoch unterschiedliche Galenik und Frequenz
(-) 0 0 (+)	Einzelfallanalyse: 60% 33% 31% 8%	17% 54% 30% 37% 5%	% Verminderung Kopfschmerzstage: 11% Komplexe Studie, verwirrende Beschreibung von Drop-outs, schwer interpretierbar
0	k.A.	Attackenrückgang von 4,3 auf 2,05 vs. 4,4 auf 2,15. Intensitätsrückgang nur mit Flunarizin	Unklare Drop-out-Rate. Sehr hohe und schnelle Ansprechraten

6.2 Tabellen zu den Studien zu Hypericumextrakten bei Depression

Erklärung der Darstellungsweise und Abkürzungen

Autor	Zahl rand./ anal.	Diagnose Depression (Klassifikation)	%weiblich/ Alter/ Dauer Epis.	Conceal./ Jadad score	Ther.dauer Zentrenzahl Zentrenart	Präparat Extrakt-dosis Hypericindos.	Kontrolle(n)
-------	-------------------------	---	-------------------------------------	-----------------------------	---	--	--------------

Spalte 1 Autor: Erstautor und Publikationsjahr (bei Jahren mit 19hundert nur die letzten 2 Ziffern)

Spalte 2: Zahl rand. = Anzahl randomisierter Patienten, anal. = Zahl in die Auswertung einbezogener Patienten

Spalte 3 Diagnose Depression (Klassifikation): In der Publikation genannte Diagnosen der Studienpatienten (Klassifikationssystem)

Spalte 4: %weiblich = Anteil der weiblichen Patienten in Prozent, Alter = Alter in Jahren (Mittelwert oder Bereich je nach Verfügbarkeit), Dauer Epis. = Bestehensdauer der aktuellen depressiven Symptomatik

Spalte 5: Conceal. = concealment, d.h. die Beurteilung der Methode, mit der sichergestellt wurde, dass der einschließende Arzt vor Aufnahme eines Patienten in die Studie nicht wußte, welche Behandlung dieser bekommen würde (adäq. = Methode adäquat); Jadadscore = Punktzahl bei den einzelnen Items des Jadadscores (Item 1 maximal 2 Punkte für Randomisation, Item 2 maximal 2 Punkte für Doppelblindbedingungen, Item 3 maximal 1 Punkt zur Beschreibung von Studienabbrüchen und Ausschlüssen)

Spalte 6: Ther.dauer = Dauer der Behandlungsphase, Zentrenzahl = Zahl der Prüfzentren, Zentrenart = Art der Prüfzentren

Spalte 7 Präparat, tägliche Extrakt-dosis und tägliche Gesamthypericindosis

Spalte 8 Kontrolle(n): Art und ggfs. Dosis der Vergleichsintervention

Vote count	Responder	Andere Ergebnisse	Bemerkungen
---------------	-----------	-------------------	-------------

Spalte 9 Vote count: Grobkategorisierung der Ergebnisse in + = Prüfbehandlung signifikant besser als Kontrolle, (+) = Trend zugunsten der Prüfbehandlung, 0 = kein Unterschied, (-) = Trend zugunsten der Kontrollbehandlung, - = Kontrollbehandlung signifikant besser

Spalte 10 Responder: Therapieerfolge in Prüf- vs. Kontrollgruppe (in Prozent oder als Anzahl Therapieerfolge/Anzahl Behandelte)

Spalte 11 Andere Ergebnisse: Kurzzusammenfassung anderer Ergebnisse

Spalte 12 Bemerkungen

Tabelle A 11:
Randomisierte Studien zu Hypericumextrakten vs. Placebo zur Behandlung von Depressionen

Autor	Zahl rand./ anal.	Diagnose Depression (Klassifikation)	%weiblich/ Alter/ Dauer Epis.	Conceal./ Jadad score	Ther.dauer Zentrenzahl, -art (Land)	Präparat Extrakt-dosis Hypericindos.	Kontrolle(n)
Bjerkenstedt 2000 (zitiert nach (338))	170 163	major (leicht bis mittelschwer, DSM IV 296.31/32)	k.A. k.A. k.A.	k.A. 1 – 1 – 0	4-6 w 15 Praxen (S)	LI 160 900 mg k.A.	Placebo Fluoxetin 20 mg
Halama 91 (159;160)	50 50	neurotische D., kurz- dauernde depressive Verstimmung (ICD 9)	64% 47 k.A.	adäquat 2 – 2 – 1	4 w 1 psychiatr. Praxis (D)	LI 160 900 mg 1,08 mg	Placebo
Hänsen 96 & 93 (173;174)	108 101	major (leicht bis mittelschwer, DSM III R)	66% 52 2 w – 6 m	adäquat* 2 – 2 – 1	4 w 11 psychiatr./ allgmed. Pr. (D)	LI 160 900 mg 2,7 mg	Placebo
Harrer 91 (166)	120 94	neurotische D., kurz- dauernde depressive Verstimmung (ICD 9)	61% 49 k.A.	k.A. 1 – 1 – 0	6 w 6 internistische Praxen (D)	Psychotonin M, 500 mg 0,75 mg	Placebo
Hoffmann 79 (186)	60 60	9 leicht, 45 mittel, 6 schwer (keine Kriterien)	48% 49 k.A.	k.A. 1 – 1 – 1	6 w k.A.	Hyperforat k.A. 0,6 mg	Placebo
Hübner 93 & 94 (196)	40 39	leichte neurotische D., kurz-dauernde d. Verstimmung (ICD 9)	55% 51 k.A.	adäquat 2 – 2 – 1	4 w 1 internistische Praxis (D)	LI 160 900 mg 2,7 mg	Placebo
König (250)	112 81	psychoaffektive Störung mit depres- siver Stimmung	75% 45 k.A.	adäquat 2 – 2 – 1	6 w 50 Allgemein- med. (CH)	Z 90017 500-1000 mg 1-2 mg	Placebo
Laakmann 98 (258;259)	147 147	major (Episode leicht bis mittel, DSM IV)	80% 53 k.A.	adäquat 2 – 2 – 1	6 w 11 unklar (D)	WS5572 900 mg k.A.	Placebo, Hypericum WS5573 (geringerer Hyperforingehalt)
Lehrl 93 (273)	50 49	neurotische D., kurz- dauernde d. Verstim- mung (major*)	82% 49 3 m	adäquat* 1 – 1 – 1	4 w 4 Praxen (D)	LI 160 900 mg 1,08 mg	Placebo
Montgomery 2000 (335)	247 247	leicht bis mittel (DSM IV 296.2x/3x)	74% 43 >2w	k.A. 1 – 2 – 0	12 w 18 Praxen (UK)	LI 160 900 mg k.A.	Placebo
Osterheider 92 (354)	46 34	mittel bis schwer (25 endogen, 12 neurotisch, 9 reaktiv)	66% 43 k.A.	k.A. 1 – 2 – 0	8 w k.A. k.A.	Psychotonin M, 500 mg 0,75 mg	Placebo
Quandt 93 (381)	88 88	leicht bis mittel (ICD 9 300.4)	73% 43 k.A.	k.A. 1 – 2 – 1	4 w 4 Praxen (D)	Psychotonin M, 500 mg 0,75 mg	Placebo
Philipp 00 (365)	240 240	mittel (ICD 10 F32.1/F33.1)	75% 47 4,5 m	adäquat 2 – 2 – 1	8 w 18 Allgemein- mediziner (D)	STEI 300 1050 mg k.A.	Placebo Imipramin 100 mg
Reh 92 (385)	50 50	neurotische D., kurz- dauernde depressive Verstimmung (ICD 9)	78% 48 k.A.	adäquat* 1 – 1 – 0	8 w 1 psychiatr. Praxis (D)	Neuroplant 500 mg 1 mg	Placebo

HAMD = Hamilton Rating Scale for Depression, MADRS = Montgomery Asberg Depression Rating Scale, SDS = Self Rating Depression Scale, CGI = Clinical Global Impression Index, D-S = Depressionsskala (von Zerssen), Bf-S = Befindlichkeitsskala (von Zerssen), B-L = Beschwerdeliste (von Zerssen), NW = Nebenwirkungen, *Beurteilung aufgrund von Informationen von Autor und/oder Sponsor

Vote count	Responder	Andere Ergebnisse	Bemerkungen
(+)	33% vs. 16% Placebo vs. 30% Fluoxetin (HAMD)	Rückgang des HAMD-Scores von 26,2 auf 15,6 vs. 26,3 auf 16,3 (Placebo) vs. 25,0 auf 15,9 (Fluoxetin)	Nur als Bericht von Kongress verfügbar, daher nur eingeschränkt beurteilbar
+	50% vs. 0% (HAMD)	Rückgang des HAMD-Scores von 18, 3 auf 10,9 vs. 18,0 auf 18,2	Keinerlei Verbesserung in der Placebogruppe irritierend
+	70% vs. 24% (HAMD)	Rückgang des HAMD-Scores von 21,0 auf 8,9 vs. 20,4 auf 14,4, D-S von 21,2 auf 9,3 vs. 19,6 auf 14,6	Gute Studie – 1993 Publikation von Zwischenergebnissen; 1996er Publikation ohne Hinweis auf diese Vorpublikation
+	66% vs. 25% (HAMD)	Rückgang des HAMD-Scores von 21,6 auf 8,9 vs. 20,9 auf 16,1, D-S von 11,7 auf 4,9 vs. 12,7 auf 8,0	Ältere, mäßig berichtete Studie mit unklaren Angaben zu Abbrüchen und Ausschlüssen, nur geringe Besserung unter Placebo
+	63% vs. 10% (Globalurteil)	Nur Daten zu einer Vielzahl von Einzelsymptomen	Ältere Studie
+	70% vs. 45% (HAMD)	HAMD-Rückgang von 12,6 auf 5,6 vs. 12,4 auf 10,3, B-L von 15,6 auf 6,8 vs. 16,7 auf 16,9	Kleine Studie mit Patienten mit sehr leichten depressiven Verstimmungen
0	53% vs. 54% (globales Patientenurteil)	Depressionsscore von 64 auf 48 vs. 65 auf 51	Nur als Dissertation verfügbar, sehr viele Studienzentren mit sehr wenigen Patienten, kein validiertes Beurteilungsinstrument
+	49% vs. 33% Placebo vs. 39% WS5573	Rückgang des HAMD-Scores von 20,9 auf 10,7 vs. 21,2 auf 13,3 Placebo und 20,3 auf 11,8 unter WS5573	Präparat mit geringerem Hyperforingehalt nicht signifikant wirksamer als Placebo
0	16% vs. 8% (HAMD)	Rückgang des HAMD-Scores von 23,7 auf 17,4 vs. 21,6 auf 16,8	Niedrige Ansprechraten in beiden Gruppen
0	47% vs. 47% (HAMD – aus Graphik extrapoliert)	Keine signifikanten Unterschiede in HAMD und MADRS-Scores sowie CGI Beurteilung	Nur als Abstract und Poster verfügbar – nur eingeschränkt beurteilbar
0	keine Responder (HAMD) in beiden Gruppen	Rückgang des HAMD-Scores von 22,4 auf 17,1 vs. 22,1 auf 16,8	Nur als Abstract publiziert, zusätzliche Informationen aus unvollständigem Auswertungsbericht
+	66% vs. 7% (HAMD)	Rückgang des HAMD-Scores von 17,8 auf 5,2 vs. 17,3 auf 15,5	Auffällig geringes Ansprechen unter Placebo
+	76% vs. 63% Placebo vs. 67% Imipramin (HAMD)	Rückgang des HAMD-Scores um $15,4 \pm 8,1$ Punkte vs. $12,1 \pm 7,4$ (Placebo) vs. $14,2 \pm 7,3$ (Imipramin); ähnlich bei SDS und Lebensqualität	Gut berichtete, sorgfältige Studie. Hohe Responderate in der Placebogruppe
+	80% vs. 44% (HAMD)	Rückgang des HAMD-Scores von 19,0 auf 6,7 vs. 20,0 auf 11,4, D-S von 25,7 auf 9,5 vs. 30,6 auf 19,2	Zusatzinformation zu Methoden und Ergebnissen deuten auf sorgfältige unizentrische Studie

Tabelle A11 (Fortsetzung)

Randomisierte Studien zu Hypericumextrakten vs. Placebo zur Behandlung depressiver Verstimmungen

Autor	Zahl rand./ anal.	Diagnose Depression (Klassifikation)	%weiblich/ Alter/ Dauer Epis.	Conceal./ Jadad score	Ther.dauer Zentrenzahl, -art (Land)	Präparat Extrakt-dosis Hypericindos.	Kontrolle(n)
Schlich 87 (396)	46 46	leicht bis mittelschwer	63% 42 k.A.	adäquat* 1 – 2 – 0	4 w 1 Praxis (D)	Psychotonin M, 350 mg 0,5 mg	Placebo
Schmidt 89 (399)	40 28	leicht bis mittel	64% 47 k.A.	adäquat* 1 – 2 – 1	4 w 1 Allg.med., 1 Internist (D)	Psychotonin M, 500 mg 0,75 mg	Placebo
Schmidt 93 (400)	65 60	neurotische D., kurz- dauernde depressive Verstimmung (ICD 9)	77% 44 k.A.	adäquat* 1 – 1 – 1	6 w 3 niedergel. Ärzte (D)	LI 160 900 mg 1,08 mg	Placebo
Schrader 98 (404)	162 159	leicht bis mittel (ICD 10 F32.0/F32.1)	67% 43 k.A.	adäquat 2 – 2 – 1	6 w 16 niedergel. Ärzte (D)	ZE 117 500 mg 1 mg	Placebo
Shelton 2001 (407)	200 195	major (DSM IV)	64% 42 28 m	adäquat 2 – 2 – 1	8 w 11 Uni-Ambu- lanzen (USA)	LI 160 1200 mg k.A.	Placebo
Sommer 93 (419) & Harrer 93/94 (168;169)	105 99	neurotische D., kurz- dauernde depressive Verstimmung (ICD 9)	72% 45 k.A.	adäquat 2 – 2 – 1	4 w 3 niedergel. Ärzte (D)	LI 160 900 mg 1,08 mg	Placebo
Volz 2000 (484)	142 140	major (Episode leicht bis mittel, DSM IV)	80% 47 8 m	k.A. 1 – 2 – 0	6 w 17 versch. Praxen (D)	D-0496 500 mg 1 mg	Placebo
Winkel 2000 (515;516)	119 95	leicht bis mittel (ICD 10, DSM III-R) bei Alkoholentzugspat.	37% 43 k.A.	k.A. 1 – 1 – 1	6 w 1 Entzugskl. (D)	LI 160 900 mg k.A.	Placebo
Witte 95 (518)	97 88	mittel (ICD 10 F 32.1)	66% 43 k.A.	k.A. 1 – 2 – 1	6 w 5 Allgmed. (D)	Psychotonin forte, 240 mg, 2,4 mg	Placebo

Vote count	Responder	Andere Ergebnisse	Bemerkungen
+	68% vs. 13% (HAMD)	Rückgang des HAMD-Scores von 22,9 auf 16,4 vs. 24,0 auf 29,6	Ältere Studie mit zum Teil widersprüchlichen Angaben (z.B. hohe HAMD-Scores bei leichter bis mittelschwerer Symptomatik)
+	50% vs. 20% (HAMD)	Rückgang des HAMD-Scores von 29,3 auf 9,8 vs. 29,5 auf 19,5	Ältere Studie, hohe HAMD-Werte passen nicht zur Diagnose leichte bis mittelschwere Depression
(+)	63% vs. 18% (HAMD)	Rückgang des HAMD-Scores von 16,6 auf 7,3 vs. 16,4 auf 8,2	Nur geringe Unterschiede bei den HAMD-Scorewerten, jedoch deutliche bzgl. Responderkriterium
+	56% vs. 15% (HAMD)	Rückgang des HAMD-Scores von 20,1 auf 10,5 vs. 18,7 auf 17,9, Besserung bei VAS-Urteil Patienten 68 mm vs. 44 mm	Sorgfältige Studie, auffällig geringe Besserung in der Placebogruppe
0 – (+)	27% vs. 19% (HAMD)	Keine Unterschiede bei HAMD-Score Rückgang, BDI-Rückgang von 22,0 auf 16,0 vs. 24,2 auf 18,7	Methodisch hochwertige Studie; ungewöhnlich langes Bestehen der aktuellen depressiven Episode deutet auf ungewöhnliche Stichprobe
+	54% vs. 25% (HAMD)	Rückgang des HAMD-Scores von 15,8 auf 7,2 vs. 15,8 auf 11,3	Mehrfach publizierte Studie
+	67% vs. 52% (CGI) bzw. 66% vs. 55% (globales Patientenurteil)	Rückgang des HAMD-Scores von 21,0 auf 12,0 vs. 20,7 auf 14,3; Rückgang D-S von 18,4 auf 10,0 vs. 18,0 vs. 12,6	Sorgfältige Studie
+	58,5% vs. 36,5% (Globalbeurteilung der Wirksamkeit)	Rückgang des HAMD-Scores von 16,6 auf 6,3 vs. von 16,7 auf 9,3	Nur als Abstract und kurzes Buchkapitel verfügbar, daher nur eingeschränkt beurteilbar
+	79% vs. 56% (HAMD)	Rückgang des HAMD-Scores von 24,6 auf 7,9 vs. 22,7 auf 10,4, D-S von 25,0 auf 7,0 vs. 21,4 auf 10,9	Sorgfältige Studie

Tabelle A12:
Randomisierte Studien zu Hypericumextrakten vs. andere Medikamente zur Behandlung von Depressionen

Autor	n rand. n anal.	Diagnose Depression (Klassifikation)	%weiblich/ Alter/ Erkr.dauer	Conceal./ Jadad score	Ther.dauer Zentrenzahl, -art (Land)	Präparat Extrakt-dosis Hypericindos.	Kontrolle(n)
Bergmann 93 (38)	80 76	leicht bis mittel (ICD 10 F32.0/1, F33.0/1)	70% 55 5,5 m	k.A. 1 – 1 – 1	6 w 1 psychiatr. Praxis (D)	Esbericum k.A. 0,75 mg	Amitriptylin 30 mg
Bjerkenstedt 2000 (338)	170 163	leicht bis mittel (DSM IV 296.31/32)	k.A. k.A. k.A.	k.A. 1 – 1 – 0	4-6 w 15 Praxen (S)	Li 160 900 mg k.A.	Fluoxetin 20 mg Placebo
Brenner 2000 (50)	30 28	major (DSM IV, erste oder rezidivierend)	63% 21-65 k.A.	k.A. 1 – 2 – 0	7 w 1 k.A. (USA)	LI 160 900 mg k.A.	Sertralin 75 mg
Harrer 93 & 94 (164;165)	102 86	einzelne, mittelgradig (ICD 10 F32.1)	72% 46 > 2 w	adäquat 2 – 2 – 1	4 w 6 psychiatr. Praxen (D)	LI 160 900 mg 2,7 mg	Maprotilin 75 mg
Harrer 99 (167)	161 149	leicht bis mittel (ICD10 F32.0, F32.1, erste Episode)	87% 69 k.A.	k.A. 1 – 1 – 1	6 w 17 Praxen (D)	LoHyp 800 mg k.A.	Fluoxetin 20 mg
Kugler 90 (254)	80 78	leicht bis mittel (endogene D. ausgeschlossen)	61% 51 k.A.	adäquat* 1 – 0 – 1	4 w 1 Forschungs- zentrum (D)	Psychotonin M, 500 mg 0,75 mg	Bromazepam 6 mg
Philipp 2000 (365)	240 240	mittel (ICD 10 F32.1/F33.1)	75% 47 4,5 m	adäquat 2 – 2 – 1	8 w 18 Allgemein (D)	STEI 300 1050 mg k.A.	Imipramin 100 mg Placebo
Schrader 2000 (403)	240 238	leicht bis mittel (ICD 10 F32.0, F32.1)	65% 46 k.A.	k.A. 1 – 2 – 1	6 w 7 Internisten (D)	Ze 117 500 mg k.A.	Fluoxetin 20 mg
Vorbach 93 (487) & 94 (488)	135 130	major (DSM III R)	47% 53 k.A.	adäquat 2 – 2 – 1	6 w 20 versch. (D)	LI 160 900 mg 2,7 mg	Imipramin 75 mg
Vorbach 97 (486)	209 186	schwer (ICD 10 F 33.2)	74% 49 k.A.	adäquat 2 – 2 – 1	6 w 20 psychiatr. Praxen (D)	LI 160 1800 mg 5,4 mg	Imipramin 150 mg
Wheatley 97 (499)	165 156	major (DSM IV)	76% 40 k.A.	adäquat 2 – 2 – 0	6 w 18 GP und 1 Ambulanz (GB)	LI 160 900 mg 0,96 mg	Amitriptylin 75 mg
Woelk (520)	324 324	leicht bis mittel (ICD 10 F32.0/1, F33.0/1)	71% 46 k.A.	k.A. 2 – 2 – 1	6 w 20 (Inn., Allg- med, Psych, D)	ZE 117 500 mg 1 mg	Imipramin 150 mg

Legende siehe Tabelle A 11

Vote count	Responder	Andere Ergebnisse	Bemerkungen
0	80% vs. 70% (HAMD)	Rückgang des HAMD-Scores von 15,8 auf 6,3 vs. 15,3 auf 6,7; Bf-S von 31,5 auf 24,2 vs. 28,2 auf 24,1	Sehr niedrige Amitriptylindosis, HAMD-Baselinewerte deuten auf relativ leichte Symptomatik
0	33% vs. 30% Fluoxetin vs. 16% Placebo (HAMD)	Rückgang des HAMD-Scores von 26,2 auf 15,6 vs. 25,0 auf 15,9 (Fluoxetin) vs. 26,3 auf 16,3 (Placebo)	Nur als Bericht von Kongress verfügbar, daher nur eingeschränkt beurteilbar
0	7/13 vs. 6/15	Rückgang des HAMD-Scores von 21,3 auf 12,7 vs. 21,7 auf 12,5, bei D-S von 24,0 auf 18,4 vs. 15,9 auf 10,5	Kleine Studie mit erheblichem und ungleichem Drop-out (7/15 vs. 3/15), allerdings intent-to-treat ausgewertet
0	53% vs. 55% (HAMD)	Rückgang des HAMD-Scores von 20,5 auf 11,2 vs. 21,5 auf 10,5, D-S von 25,6 auf 16,1 vs. 25,2 auf 13,9	Gut berichtete Studie, eng begrenzte Diagnose
0	50/70 vs. 57/79 (HAMD)	Rückgang des HAMD-Scores von 16,6 auf 7,9 vs. 17,2 auf 8,1; Ergebnisse ähnlich für SDS und CGI, NW in 12 vs. 17 Patienten	Studie in älteren Patienten
0	k.A.	Rückgang SDS-Scorewerte von 49 auf 40 vs. 49 auf 40, DSI von 45 auf 35 vs. 44 auf 35	Ältere Studie, keine Blindbedingungen, Bromazepam problematische Vergleichssubstanz
+	76% vs. 63% Placebo vs. 67% Imipramin (HAMD)	Rückgang des HAMD-Scores um $15,4 \pm 8,1$ Punkte vs. $12,1 \pm 7,4$ (Placebo) vs. $14,2 \pm 7,3$ (Imipramin); ähnlich bei SDS und Lebensqualität	Gut berichtete, sorgfältige Studie. Hohe Responderate in der Placebogruppe
(+)	60% vs. 40% (HAMD)	Rückgang des HAMD-Scores von 19,6 auf 11,5 vs. 19,5 auf 12,2 ($p = 0,09$), Besserung bei VAS-Urteil Patienten 32 mm vs. 28 mm	Sorgfältige Studie, nur 2 Drop-outs
0	63% vs. 54% (HAMD)	Rückgang des HAMD-Scores von 20,2 auf 8,8 vs. 19,4 auf 10,7, D-S von 39,6 auf 27,2 vs. 39,0 auf 29,2	Sorgfältige Studie
0/(-)	35% vs. 41%	Rückgang des HAMD-Scores von 25,3 auf 14,4 vs. 26,1 auf 13,4, D-S von 28,9 auf 16,5 vs. 28,5 auf 13,6	Einzige Studie explizit zu schwerer Depression; gute Qualität, Imipramin marginal effektiver, aber mehr Nebenwirkungen; hohe Hyper.dosis
(-)	60% vs. 78% (HAMD)	Rückgang des HAMD-Scores von 20 auf 9 vs. 21 auf 5, MADRS von 27 auf 10 vs. 26 auf 6	Studie aus GB mit leichter Überlegenheit von Amitriptylin, aber deutlich mehr Nebenwirkungen
0	43% vs. 40% (HAMD)	Rückgang des HAMD-Scoren um $12,4 \pm 5,9$ Punkte vs. $12,9 \pm 6,8$ Punkte; Patientenbewertung 2,44 vs. 2,16 (7-stufige Skala, niedrige Werte besser)	Sorgfältige Studie, Daten zu intent-to-treat wie auch zur per-protocol Analyse. Hohe Imipramindosis führte zu vielen Abbrüchen

7 Erläuterungen zum Verständnis der Effektmaßdarstellungen

Die folgenden Erläuterungen sollen Lesern, die in der Interpretation von Effektmaßen und deren graphischer Darstellung nicht geübt sind, das Verständnis der mithilfe der Cochrane Collaboration Review Manager Software erzeugten Abbildungen bzw. Darstellungen erleichtern.

1. Dichotome Maße

Dichotome Effektmaße (d.h. Effektmaße zu ja/nein-Daten) werden in der vorliegenden Monographie als relative Risiken bzw. als Reponderratio (RR) dargestellt. Die Berechnung wird anhand der ersten Studie in Abbildung 1 (Borgesen) erklärt. In dieser Studie waren 14 von 30 in der Verumgruppe (Spalte Treatment) behandelten Patienten Responder. Der Responderanteil in der Verumgruppe beträgt also $14/30 = 46,7\%$. In der Placebogruppe (Spalte Control) waren 9 von 30 Patienten Responder, dementsprechend war der Responderanteil $9/30 = 30\%$. Zur Ermittlung der Responderratio wird der Responderanteil in der Verumgruppe (46,7%) durch den Responderanteil in der Placebogruppe (30%) dividiert. Dies ergibt den Wert 1,56. Dieser Wert ist in der graphischen Darstellung durch das Quadrat wiedergegeben und in der letzten Spalte als Zahl angegeben.

Comparison: 01 Propranolol vs. placebo

Outcome: 04 responder including cross-over trials

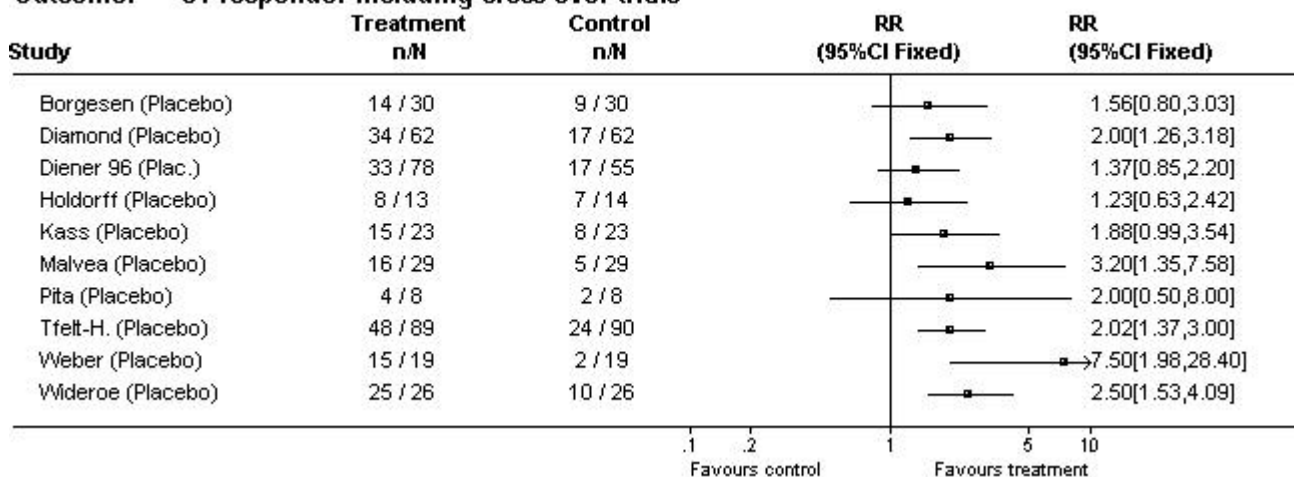


Abbildung. 1:

Responderratio in Studien Propranolol vs. Placebo

(Abbildung aus Cochrane Collaboration Review Manager Software, Version 4.1)

n = Anzahl Responder, N = Anzahl Patienten in der jeweiligen Gruppe, RR = Responderratio, 95%CI fixed = 95% Konfidenzintervall

Insbesondere die graphische Darstellung erlaubt eine intuitive Erfassung der gesamten Studienergebnisse. Für jede Studie wird nach dem obengenannten Vorgehen die RR berechnet. Bei Studien, bei denen unter Verum mehr Patienten Responder sind als unter Placebo, ergibt sich immer eine RR größer als 1. Sind unter Verum und Placebo die Responderanteile gleich (z.B. jeweils 50%), so ist die $RR = 50\%/50\% = 1$. Eine $RR = 1$ steht also für gleiche Ergebnisse in beiden Gruppen; der Wert ist durch die senkrechte Linie in der Graphik markiert. Ist Verum Placebo unterlegen, ergibt sich ein Wert unter 1, z.B. bei einer Responderrate von 25% unter Verum und 50% unter Placebo der Wert 0,5.

Die waagrechten Linien um das Quadrat stellen das 95% Konfidenzintervall der jeweiligen RR dar. Berührt die waagrechte Linie die senkrechte Linie nicht (schließt das Konfidenzintervall also den Wert 1 nicht ein) ist der Unterschied zwischen Verum (Treatment) und Placebo (Control) statistisch signifikant ($p < 0,05$).

So lässt sich aus Abbildung 1 sofort erkennen, dass in allen Studien Propranolol zumindest tendenziell einer Placebobehandlung überlegen, der Unterschied aber nur in 5 Studien (Diamond, Malvea, Tfelt-Hansen, Weber und Wideroe) statistisch signifikant war.

Allerdings ist zu beachten, dass es vom Zielparameter abhängt, ob im jeweiligen Fall Werte größer als 1 eine Überlegenheit oder eine Unterlegenheit darstellen. Ist der Zielparameter ein positives Ereignis (z.B. Response) spricht man von einer Responderratio und Werte größer 1 sprechen für eine Überlegenheit der Verum- bzw. Testgruppe. Ist der Zielparameter jedoch ein negatives Ereignis (z.B. Tod) spricht man von einem relativen Risiko und Werte größer 1 stehen für ein schlechteres Ergebnis in der Verum- bzw. Testgruppe. So werden in der Abbildung 4 z.B. die in den einzelnen Studien beobachteten relativen Risiken des Auftretens von Nebenwirkungen unter Placebo und anderen Substanzen verglichen. In der ersten Studie von Albers traten bei 15 von 20 Propranololpatienten (75%) Nebenwirkungen auf, unter Flunarizin dagegen bei 20 von 20 Patienten (100%). Daraus ergibt sich eine RR von 0,75 (95%-Konfidenzintervall 0,58-0,97), d.h. Nebenwirkungen waren unter Propranolol signifikant seltener.

Comparison: 02 Propranolol vs. other drugs

Outcome: 09 number of patients with side effects including cross-over trials

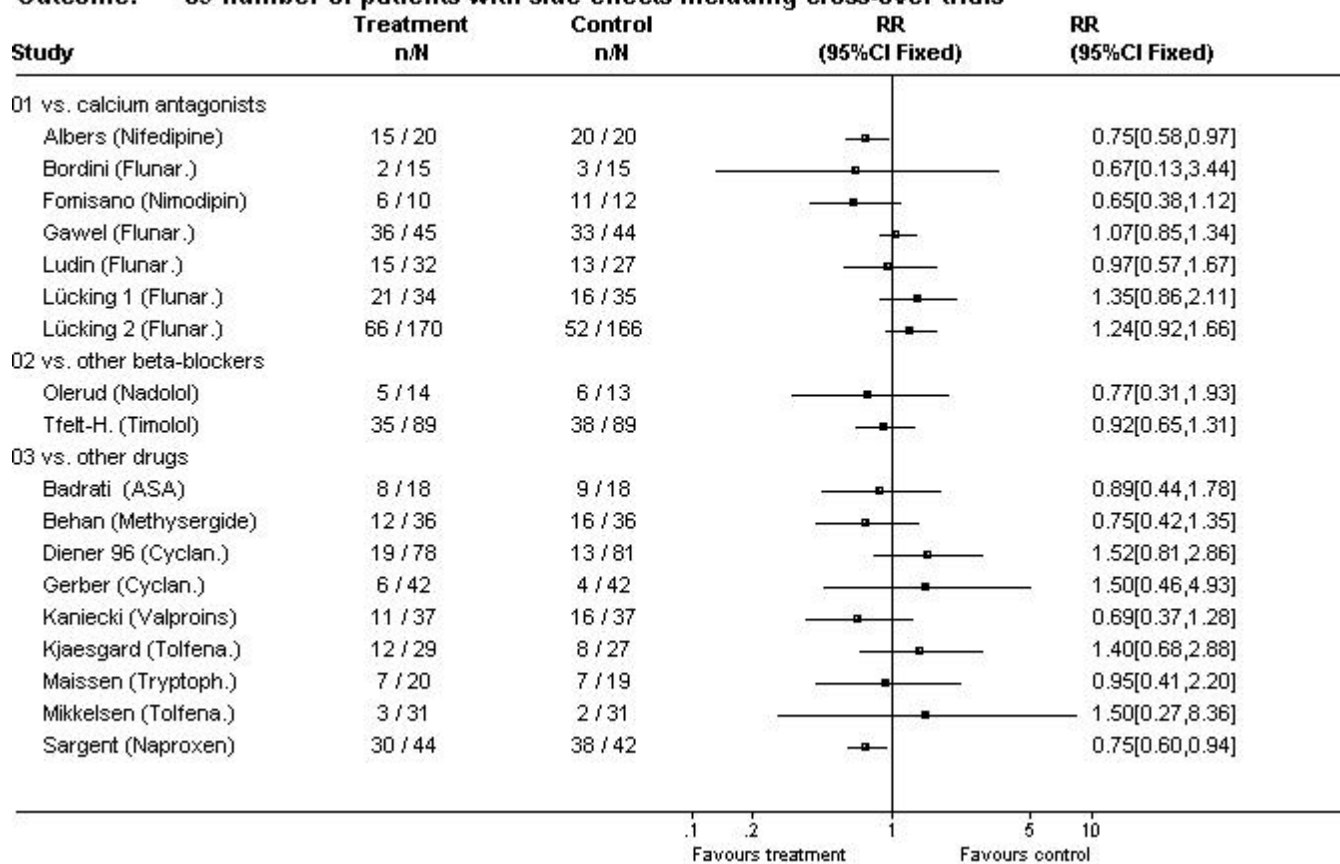


Abbildung 2:

Relative Risiken bzgl. des Auftretens von Nebenwirkungen in Studien Propranolol vs. andere Substanzen
(Abbildung aus Cochrane Collaboration Review Manager Software, Version 4.1)

n = Anzahl Responder, N = Anzahl Patienten in der jeweiligen Gruppe, RR = relatives Risiko, 95%CI fixed = 95% Konfidenzintervall

2. Intervall- und kontinuierliche Maße

Die Ergebnisse zu Parametern mit Intervall- oder kontinuierlicher Skalierung (z.B. Anzahl der Kopfschmerztag, Intensitätsscores) werden in der vorliegenden Monographie als standardisierte Mittelwertsdifferenzen dargestellt. Dabei wird die Differenz zwischen den Mittelwerten der Test- und Kontrollgruppe durch die „gepoolte“ Standardabweichung (d.h. es wird eine „gemeinsame“ Standardabweichung beider Gruppen berechnet) geteilt. Dieses Maß hat den Vorteil, dass mit verschiedenen Methoden gemessene Parameter auf dasselbe Darstellungsformat gebracht werden können, hat aber den Nachteil, dass es nicht intuitiv verständlich ist. Allgemein werden standardisierte Mittelwertsdifferenzen von bis zu 0,3 (Standardabweichungen) als kleine Effekte interpretiert, Werte von 0,4 bis 0,6 als mittlere Effekte und Werte von 0,7 und mehr als große Effekte. Jedoch kann auch eine Studie mit geringer Mittelwertsdifferenz bei kleiner Standardabweichung eine große standardisierte Mittelwertsdifferenz haben.

In Abbildung 3 werden z.B. die standardisierten Mittelwertsdifferenzen (SMD) für Kopfschmerz-Häufigkeitsmaße verwendet. Seltener Kopfschmerzen sind erwünscht, eine negative

Mittelwertsdifferenz steht daher für eine Überlegenheit der Testgruppe. In der Studie von Albers traten in der Testgruppe (Propranolol) häufiger Kopfschmerzen auf als in der mit Nifedipin behandelten Kontrollgruppe. Der SMD-Wert beträgt hier 0,80 Standardabweichungen, das entsprechende 95%-Konfidenzintervall ist –0,16 bis 1,76 (der Unterschied schließt den Wert 1 ein, ist also nicht signifikant). Wie in Abbildung 3 schnell zu erkennen ist, ergab sich nur bei einer einzigen Studien ein signifikanter Unterschied (Kjaesgard).

Comparison: 02 Propranolol vs. other drugs

Outcome: 05 frequency measures including cross-over trials

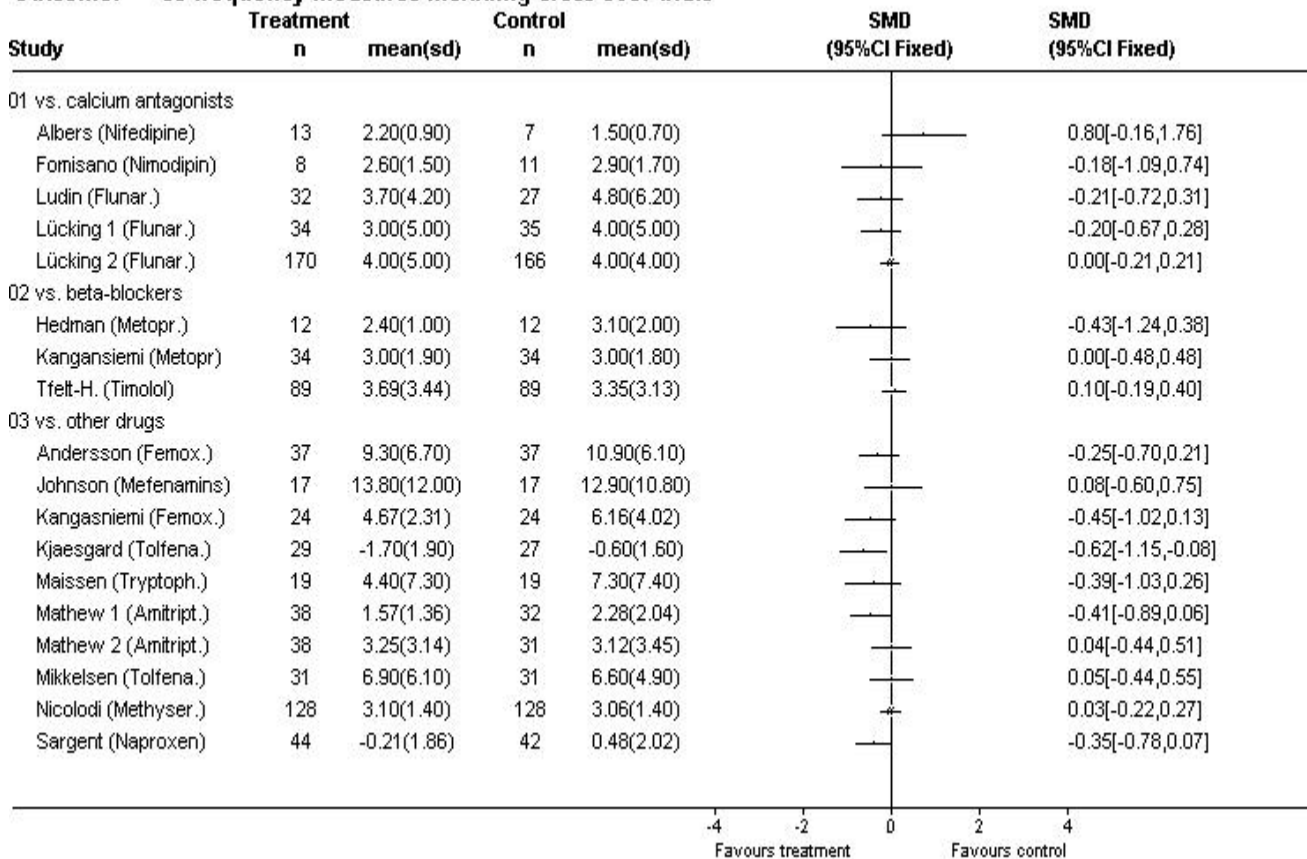


Abbildung. 3:

Standardisierte Effektgrößen zu Häufigkeitsmaßen in Studien Propranolol vs. andere Substanzen

(Abbildung aus Cochrane Collaboration Review Manager Software, Version 4.1)

n = Anzahl Patienten in der jeweiligen Gruppe, mean (sd) = Mittelwert (Standardabweichung), SMD = standardisierte Mittelwertsdifferenz, 95%CI fixed = 95%-Konfidenzintervall

8 Literaturverzeichnis

- (1) ANONYMOUS. Propranolol for prevention of migraine headaches. *Med Lett Drugs Ther* 1979; **21**: 77-78.
- (2) ANONYMOUS. Cochrane Reviewers Handbook 4.1.3 (updated June 2001). Clarke M, Oxman AD, Herausgeber. The Cochrane Library, issue 3, 2001. Oxford: Update Software.
- (3) AGNOLI A, BUSSONE G, MAILLAND F, MANZONI GCMN, NAPPI G. Dihydroergokryptine vs flunarizine in the basic treatment of migraine without aura. *Cephalalgia* 1991; **11 (suppl 11)**: 216-217.
- (4) AGRAWAL A, DIXIT SP, DUBEY GP, GUPTA BS. Clinical evaluation of anti-depressant properties of basant (*Hypericum perforatum*). *Pharmacopsychocologia* 1994; **7**: 253-256.
- (5) AHONEN E, HAKUMÄKI M, MAHLAMÄKI S ET AL. Effectiveness of acupuncture and physiotherapy on myogenic headache: a comparative study. *Acupunct Electrother Res Int J* 1984; **9**: 141-150.
- (6) AHUJA GK, VERMA AK. Propranolol in prophylaxis of migraine. *Indian J Med Res* 1985; **82**: 263-265.
- (7) AIKINS-MURPHY P. Alternative therapies for nausea and vomiting of pregnancy. *Obstet Gynecol* 1998; **91**: 149-155.
- (8) AL-DEEB SM, BIARY N, BAHOU Y, AL-JABERI M, KHOJA W. Flunarizine in migraine: a double-blind placebo-controlled study (in a Saudi population). *Headache* 1992; **32**: 461-462.
- (9) AL-QASSAB HK, FINDLEY LJ. Comparison of propranolol LA 80 mg and propranolol LA 160 mg in migraine prophylaxis: a placebo controlled study. *Cephalalgia* 1993; **13**: 128-131.
- (10) ALBERS GW, SIMON LT, HAMIK A, PEROUTKA SJ. Nifedipine versus propranolol for the initial prophylaxis of migraine. *Headache* 1989; **29**: 215-218.
- (11) ALBERTINI H. Evaluation d'un traitement homéopathique de la névralgie dentaire. In: Boiron J, Belon P, Hariveau E, Hrsg. Recherche en homéopathie. Lyon: Fondation française pour la recherche en homéopathie, 1986: 75-77.
- (12) ALLAIS G, DE LORENZO C, QUIRICO PE, LUPI G, CIOCHETTO D, BENEDETTO C. Acupuncture versus flunarizine in the prophylactic treatment of migraine without aura: A randomized controlled trial. *Cephalalgia* 1997; **17**: 442-443.

- (13) AMERY WK. Flunarizine, a calcium channel blocker: a new prophylactic drug in migraine. *Headache* 1983; **23**: 70-74.
- (14) AMERY WK. Onset of action of various migraine prophylactics. *Cephalalgia* 1988; **8 (suppl 8)**: 11-13.
- (15) AMERY WK, CAERS LI, AERTS TJ. Flunarizine, a calcium entry blocker in migraine prophylaxis. *Headache* 1985; **25**: 249-254.
- (16) ANDERSSON PG. Flunarizine and propranolol in the treatment of migraine - a double-blind trial. *J Neurol* 1985; **232 (suppl)**: 220.
- (17) ANDERSSON PG, DAHL S, HANSEN JH ET AL. Prophylactic treatment of classical and non-classical migraine with metoprolol--a comparison with placebo. *Cephalalgia* 1983; **34**: 207-212.
- (18) ANDERSSON PG, PETERSEN EN. Propranolol and femoxetine, a HT-uptake inhibitor, in migraine prophylaxis. A double-blind crossover study. *Acta Neurol Scand* 1981; **64**: 280-288.
- (19) ARAUJO MS. Does the choice of placebo determine the results of clinical studies on acupuncture? A meta-analysis of 100 clinical trials. *Forsch Komplementärmed* 1998; **5 (suppl 1)**: 8-11.
- (20) ARMSTRONG NC, ERNST E. The treatment of eczema with Chinese herbs: a systematic review of randomized controlled trials. *Br J clin Pharmacol* 1999; **48**: 262-264.
- (21) ARZNEIMITTELKOMMISSION DER DEUTSCHEN ÄRZTESCHAFT. Handlungsleitlinie - chronische Kopfschmerzen. *Z ärztl Fortbild Quallsich* 1997; **91**: 747-750.
- (22) ASKIE L. Broad versus narrow questions: empirical evidence of lumping and splitting in Cochrane perinatal reviews. *The Cochrane Methodology Register in The Cochrane Library*, issue 3, 2001. Oxford: Update Software.
- (23) BAILAR JC. The promise and problems of meta-analysis. *N Engl J Med* 1997; **337**: 559-560.
- (24) BAISCHER W. Acupuncture in migraine: Long-term outcome and predicting factors. *Headache* 1995; **35**: 472-474.
- (25) BAKER C. Double-blind evaluation of flunarizine and placebo in the prophylactic treatment of migraine. *Headache* 1987; **27**: 288.

- (26) BALDRATI A, CORTELLI P, PROCACCIANTI G ET AL. Propranolol and acetylsalicylic acid in migraine prophylaxis. Double-blind crossover study. Acta Neurol Scand 1983; **67**: 181-186.
- (27) BANERJEE M, FINDLEY LJ. Propranolol in the treatment of acute migraine attacks. Cephalalgia 1991; **4**: 193-196.
- (28) BARNES J, RESCH KL, ERNST E. Homeopathy for postoperative ileus? A meta-analysis. J Clin Gastroenterol 1997; **25**: 628-633.
- (29) BARRETT B, VOHMANN M, CALABRESE C. Echinacea for upper respiratory infection. J Fam Pract 1999; **48**: 628-635.
- (30) BASSI P, BRUNATI L, RAPUZZI B, ALBERTI E, MANGONI A. Low dose flunarizine in the prophylaxis of migraine. Headache 1992; **32**: 390-392.
- (31) BATRA YK. Acupuncture in the treatment of migraine. Am J Acupunct 1986; **14**: 135-137.
- (32) BAUST W, STÜRTZBECHER H. Akupunkturbehandlung der Migräne im Doppelblindversuch. Med Welt 1978; **29**: 669-673.
- (33) BECK R. Die kontrollierten Effizienzstudien: ein tabellarischer Überblick. AKU 1996; **24**: 205-207.
- (34) BEGG C, CHO M, EASTWOOD S ET AL. Improving the quality of reporting of randomized controlled trials. The CONSORT statement. JAMA 1996; **276**: 637-639.
- (35) BEHAN PO, REID M. Propranolol in the treatment of migraine. Practitioner 1980; **224**: 201-204.
- (36) BENDRE VV, DHARMADHIKARI SD. Arnica montana and hypericum in dental practice. Hahnemann Gleanings 1980 (**ohne Bandangabe**): 70-72.
- (37) BENSON K, HARTZ AJ. A comparison of observational studies and randomized, controlled trials. N Engl J Med 2000; **342**: 1878-1886.
- (38) BERGMANN R, NÜßNER J, DEMLING J. Behandlung leichter bis mittelschwerer Depressionen. TW Neurol Psychiatr 1993; **7**: 235-240.
- (39) BERMAN B, EZZO J, HADHAZY V, SWYERS JP. Is acupuncture effective in the treatment of fibromyalgia? Family Practice 1999; **48**: 213-218.
- (40) BERNHARDT M, LISKE E, EBELING L. Hypericum perforatum in der Therapie leichter bis mittelschwerer Depressionen: Vergleich der antidepressiven Wirksamkeit von zwei unterschiedlichen Dosierungsschemata. Unpublizierter Bericht, 1994.

- (41) BERNIK V, MAIA E. Uso do propranolol na profilaxia da enxaqueca: Estudo duplo-cego comparando propranolol a um analgesico (acetaminofen) e placebo. *Folha Med* 1978; **77**: 501-508.
- (42) BOIVIE J, BRATTBERG G. Are there long lasting effects on migraine headache after one series of acupuncture treatments? *Am J Chin Med* 1987; **15**: 69-75.
- (43) BONO G, MANZONI GC, MARTUCCI N ET AL. Flunarizine in common migraine: Italian cooperative trial. II. Long-term follow-up. *Cephalalgia* 1985; **5 (suppl 2)**: 155-158.
- (44) BONUSO S, DI-STASIO E, MARANO E ET AL. Long-term outcome of migraine therapy: predictive value of the frontotemporal nitroglycerin test. *Neurology* 1998; **51**: 1475-1478.
- (45) BORDINI CA, ARRUDA MA, CICIARELLI MC, SPECIALI JG. Propranolol vs flunarizine vs flunarizine plus propranolol in migraine without aura prophylaxis. A double-blind trial. *Arq Neuropsiquiatr* 1997; **55(3B)**: 536-541.
- (46) BORGESEN SE. Treatment of migraine with propranolol. *Postgrad Med J* 1976; **52 (suppl 4)**: 163-165.
- (47) BORGESEN SE, NIELSEN JL, MOLLER CE. Prophylactic treatment of migraine with propranolol. A clinical trial. *Acta Neurol Scand* 1974; **50**: 651-656.
- (48) BORGESEN SE, NIELSEN JL, MOLLER CE. Propranolol in migraine. *Lancet* 1974; **II**: 58.
- (49) BOYLE P, ROBERTSON C, LOWE F, ROEHBORN C. Meta-analysis of clinical trials of Permixon in the treatment of symptomatic benign prostatic hyperplasia. *Urology* 2000; **55**: 533-539.
- (50) BRENNER R, AZBEL V, MADHUSOODANAN S, PAWLOWSKA M. Comparison of an extract of hypericum (LI 160) and sertraline in the treatment of depression: a double-blind-randomized pilot study. *Clin Ther* 2000; **22**: 411-419.
- (51) BREWINGTON V, SMITH M, LIPTON D. Acupuncture as a detoxification treatment: an analysis of controlled research. *J Alt Complement Med* 1994; **11**: 289-307.
- (52) BROCKMÖLLER J, REUM T, BAUER S, KERB R, HÜBNER WD, ROOTS I. Hypericin and pseudohypericin: pharmacokinetics and effects on photosensitivity in humans. *Pharmacopsychiatry* 1997; **30 (suppl 2)**: 94-101.
- (53) BUDEIRI D, LI WAN PO A, DORNAN JC. Is Evening Primrose Oil of value in the treatment of premenstrual syndrome? *Contr Clin Trials* 1996; **17**: 60-68.

- (54) BUSSONE G, BALDINI S, D'ANDREA G ET AL. Nimodipine versus flunarizine in common migraine: a controlled pilot trial. *Headache* 1987; **27**: 76-79.
- (55) CANEPARI C, RIVA M, ERMINIO F. Effectiveness of flunarizine in chronic headache. A double-blind comparison with placebo and amitriptyline. *Clin Ter* 1985; **115**: 463-468.
- (56) CARLSSON CPO, SJÖLUND BH. Acupuncture and subtypes of chronic pain: assessment of long-term results. *Clin J Pain* 1994; **10**: 290-295.
- (57) CARLSSON J, FAHLCRANTZ A, AUGUSTINSSON LE. Muscle tenderness in tension headache treated with acupuncture or physiotherapy. *Cephalalgia* 1990; **10**: 131-141.
- (58) CARROLL JD, REIDY M, SAVUNDRA PA, CLEAVE N, MCAINSH J. Long-acting propranolol in the prophylaxis of migraine: a comparative study of two doses. *Cephalalgia* 1990; **10**: 101-105.
- (59) CECCHERELLI F, ALTAFINI L, ROSSATO M ET AL.: Trattamento agopunturale dell'emicrania senz'aura. Studio in doppio cieco vs placebo. Proceedings XV Congresso Nazionale, AISD Associazione Italiana per lo Studio del Dolore, S. Margherita Ligure, 1992, 310-318.
- (60) CECCHERELLI F, AMBROSIO F, ALTOPIEDI A ET AL. Trattamento agopunturale delle cefalee. *Min Riflessoter Laserter* 1985; **2**: 45-59.
- (61) CECCHERELLI F, CAVALLO C, AMBROSIO F, ALTOPIEDI A, MORO-BOSCOLO E, GIRON GP. Psychologische Untersuchung bei Patienten mit chronischem Schmerz, die mit Akupunkturreflextherapie behandelt wurden. *Dtsche Ztschr Akup* 1983; **5**: 105-112.
- (62) CENTER FOR CLINICAL HEALTH POLICY RESEARCH, DUKE UNIVERSITY. Drug treatments for the prevention of migraine. Technical review 2.3. Rockville (MD): Agency for Health Care Policy and Research, 1999. Available from the National Technical Information Service, Springfield, VA 22161.
- (63) CENTONZE V, MACINAGROSSA G, ATTOLINI E ET AL. Terapia preventiva dell'emicrania: flunarizina versus verapamil. *Clin Ter* 1985; **115**: 333-339.
- (64) CERBO R, CASACCHIA M, FORMISANO R ET AL. Studio clinico in doppio cieco: flunarizina contro pizotifene in un'unica somministrazione serale in pazienti emicranici. *Riv Neurol* 1985; **55**: 139-146.
- (65) CERBO R, CASACCHIA M, FORMISANO R ET AL. Flunarizine-pizotifen single-dose double-blind cross-over trial in migraine prophylaxis. *Cephalalgia* 1986; **6**: 15-18.

- (66) CHALMERS TC, BERRIER J, SACKS HS, LEVIN H, REITMAN D, NAGALINGAM R. Meta-analysis of clinical trials as a scientific discipline: II: Replicate variability and comparison of studies that agree and disagree. *Stat Med* 1987; **6**, 733-744.
- (67) CHEN GS, HWANG YC. Therapeutic effect of acupuncture for chronic pain. *Am J Chin Med* 1977; **5**: 45-61.
- (68) CHENG ACK. The treatment of headaches employing acupuncture. *Am J Chin Med* 1975; **3**: 181-185.
- (69) CHOI PT, HALPERN SH, MALIK N, JADAD AR, TRAMER MR, WALDER B. Examining the evidence in the anesthesia literature: a critical appraisal of systematic reviews. *Anesth Analg* 2001; **92**: 700-709.
- (70) CLARKE MJ, STEWART LA. Obtaining individual patient data from randomised controlled trials. In: Egger I, Smith GD, Altman DG, Hrsg. *Systematic reviews in health care: meta-analysis in context*. London: BMJ Books, 2001: 109-121.
- (71) CLAUSIUS N. Kontrollierte klinische Studien zur Homöopathie. Eine systematische Übersichtsarbeit mit Metaanalyse. Essen: KVC Verlag - Karl und Veronica Carstens-Stiftung, 1998.
- (72) CONCATO J, SHAH N, HORWITZ RI. Randomized, controlled trials, observational studies, and the hierarchy of research designs. *N Engl J Med* 2000; **342**: 1887-1892.
- (73) COOK DJ, MULROW CD, DAVIDOFF F. Synthesis of best evidence for clinical decisions. In: Mulrow CD, Cook DJ, Hrsg. *Systematic reviews - synthesis of best evidence for health care decisions*. Philadelphia: ACP, 1998: 5-12.
- (74) COOK DJ, REEVE BK, GUYATT GH, GRIFFITH LF, HEYLAND DK, TRYBA N. Stress ulcer prophylaxis in the critically ill: resolving discordant meta-analyses. *JAMA* 1996; **275**: 308-314.
- (75) CORTELLI P, SACQUEGNA T, ALBANI F ET AL. Propranolol plasma levels and relief of migraine. *Arch Neurol* 1985; **42**: 46-48.
- (76) CUCHERAT M, HAUGH MC, GOOCH M, BIOSSEL JP. Evidence of clinical efficacy of homeopathy. A meta-analysis of clinical trials. *Eur J Clin Pharmacol* 2000; **56**: 27-33.
- (77) CUTARELLA R, SANTINI S, GIUSTI A, COLANGELO U. Prophylaxis of migraine without aura: Preliminary data with lysuride and flunarizine. *Riv Neurobiol* 1995; **41**: 741-743.
- (78) DAHLÖF C. No clearcut longterm prophylactic effect of one month treatment with propranolol in migraineurs. *Cephalalgia* 1987; **7(suppl 6)**: 459-460.

- (79) DE-BOCK GH, EELHART J, VAN-MARWIJK HW, TROMP TP, SPRINGER MP. A postmarketing study of flunarizine in migraine and vertigo. *Pharm World Sci* 1997; **19**: 269-274.
- (80) DIAMOND S, FREITAG FG. A double blind trial of flunarizine in migraine prophylaxis. *Headache Quaterly* 1993; **4**: 169-172.
- (81) DIAMOND S, KUDROW L, STEVENS J, SHAPIRO DB. Long-term study of propranolol in the treatment of migraine. *Headache* 1982; **22**: 268-271.
- (82) DIAMOND S, MEDINA JL. Double blind study of propranolol for migraine prophylaxis. *Headache* 1976; **16**: 24-27.
- (83) DIAMOND S, SOLOMON GD, FREITAG FG, MEHTA ND. Long-acting propranolol in the prophylaxis of migraine. *Headache* 1987; **27**: 70-72.
- (84) DIEHL DL. Acupuncture for gastrointestinal and hepatobiliary disorders. *J Alt Complement Med* 1999; **5**: 27-45.
- (85) DIEHM C. The role of oedema protective drugs in the treatment of chronic venous insufficiency: a review of evidence based on placebo-controlled clinical trials with regard to efficacy and tolerance. *Phlebology* 1996; **11**: 23-29.
- (86) DIENER HC. Efficacy and tolerability of flunarizine and propranolol in the prophylactic treatment of migraine. *Cephalalgia* 1999; **19**: 374.
- (87) DIENER HC, BRUNE K, GERBER WD, GÖBEL H, PFAFFENRATH V. Behandlung der Migräneattacke und Migräneprophylaxe. *Deutsches Ärzteblatt* 1997; **94**: C-2277-C-2283.
- (88) DIENER HC, FOH M, IACCARINO C ET AL. Cyclandelate in the prophylaxis of migraine: a randomized, parallel, double-blind study in comparison with placebo and propranolol. *Cephalalgia* 1996; **16**: 441-447.
- (89) DIENER HC, HARTUNG E, CHRUBASIK J ET AL. A comparative study of oral acetylsalicylic acid and metoprolol for the prophylactic treatment of migraine. A randomized, controlled, double-blind, parallel group phase III study. *Cephalalgia* 2001; **21**: 120-128.
- (90) DIENER HC, SCHOLZ E, DICHGANS J ET AL. Central effects of drugs used in migraine prophylaxis evaluated by visual evoked potentials. *Ann Neurol* 1989; **25**: 125-130.
- (91) DITTMER TLJ. Die Behandlung von psychovegetativen Störungen mit Johanniskraut-Öl (Jukunda-Rot-Öl-Kapseln). *Naturheilpraxis mit Naturmedizin* 1992; **45**: 118-122.

- (92) DITZLER K, SCHATTON W. Johanniskraut bei leichten bis mittelschweren Depressionen - Ergebnisse einer plazebo-kontrollierten Doppelblindstudie. *Heilkunst* 1992; **104**: 263-270.
- (93) DOBIE RA. A review of randomized clinical trials in tinnitus. *Laryngoscope* 1999; **109**: 1202-1211.
- (94) DOERR PH, WITTCHEN HU. Ein muskel- und gefäßorientiertes Entspannungsprogramm (SEP) zur Behandlung chronischer Migränepatienten. Eine randomisierte klinische Vergleichsstudie. *Z Psychosom Med Psychoanal* 1985; **31**: 247-266.
- (95) DOWSON DI, LEWITH GT, MACHIN D. The effects of acupuncture versus placebo in the treatment of headache. *Pain* 1985; **21**: 35-42.
- (96) DREW S, DAVIES E. Effectiveness of Ginkgo biloba in treating tinnitus: double blind, placebo controlled trial. *BMJ* 2001; **322**: 1249-1252.
- (97) DRILLISCH C, GIRKE W. Ergebnisse der Behandlung von Migräne-Patienten mit Cinnarizin und Flunarizin. *Med Welt* 1980; **31**: 1870-1872.
- (98) DUMONT L, MARDIROSOFF C, TRAMÉR M. Efficacy and harm of pharmacological prevention of acute mountain sickness: a quantitative systematic review. *BMJ* 2000; **321**: 267-272.
- (99) EGGER M, DICKERSIN K, SMITH GD. Problems and limitations in conducting systematic reviews. In: Egger M, Smith GD, Altman DG, Hrsg. *Systematic reviews in health care: meta-analysis in context*. London: BMJ Books, 2001: 43-68.
- (100) EGGER M, SMITH GD. Potentials and promise. *BMJ* 1997; **315**: 1371-1374.
- (101) EGGER M, SMITH GD. Bias in location and selection of studies. *BMJ* 1998; **316**: 61-66.
- (102) EGGER M, SMITH GD. Principles of and procedures for systematic reviews. In: Egger M, Smith GD, Hrsg. *Systematic reviews in health care: meta-analysis in context*. London: BMJ Publishing Group, 2001: 32-42.
- (103) EMERSON JD, BURDICK E, HOAGLIN DC, MOSTELLER F, CHALMERS TC. An empirical study of the possible relation of treatment differences to quality scores in controlled randomized clinical trials. *Contr Clin Trials* 1990; **11**: 339-352.
- (104) ERNST E. St. John's Wort, an anti-depressant? A systematic, criteria-based review. *Phytomedicine* 1995; **2**: 67-71.
- (105) ERNST E. Acupressure for nausea: a best evidence analysis. *J Phys Med Res* 1996; **6**: 26.

- (106) ERNST E. Ginkgo biloba in der Behandlung der Claudicatio intermittens. Fortschr Med 1996; **114**: 85-87.
- (107) ERNST E. Acupuncture as a symptomatic treatment of osteoarthritis. A systematic review. Scand J Rheumatol 1997; **26**: 444-447.
- (108) ERNST E. Acupuncture/acupressure for weight reduction? Wien klin Wschr 1997; **109**: 60-62.
- (109) ERNST E. Can allium vegetables prevent cancer? Phytomedicine 1997; **4**: 79-83.
- (110) ERNST E. Are highly dilute homoeopathic remedies placebos? Perfusion 1998; **11**: 291-292.
- (111) ERNST E. Classical homoeopathy versus conventional treatment: a systematic review. Perfusion 1999; **12**: 13-15.
- (112) ERNST E. Homeopathic prophylaxis of headaches and migraine? A systematic review. J Pain Symptom Manag 1999; **18**: 353-357.
- (113) ERNST E. Second thoughts about safety of St John's wort. Lancet 1999; **354**: 2014-2015.
- (114) ERNST E. Plants with hypoglycaemic activity in humans. Phytomedicine 2001; **4**: 73-78.
- (115) ERNST E. The efficacy of Phytodolor for the treatment of musculoskeletal pain - a systematic review of randomised trials. Natural Medicine Journal 2001 (Summer): **14-17**.
- (116) ERNST E, BARNES J. Are homoeopathic remedies effective for delayed-onset muscle soreness? A systematic review of placebo-controlled trials. Perfusion 1998; **11**: 4-8.
- (117) ERNST E, CHRUBASIK S. PHYTO-ANTI-INFLAMMATORIES. A systematic review of randomized, placebo-controlled, double-blind trials. In: Panush RS, Hrsg. Rheumatic Disease Clinics of North America. Complementary and alternative therapies for rheumatic diseases II. Orlando: W.B. Saunders Company, 2000: 13-27.
- (118) ERNST E, HUNTLEY A. Tea tree oil: a systematic review of randomised clinical trials. Forsch Komplementärmed 2000; **7**: 17-20.
- (119) ERNST E, PITTLER MH. Efficacy of homeopathic arnica. A systematic review of placebo-controlled clinical trials. Arch Surg 1998; **133**: 1187-1190.
- (120) ERNST E, PITTLER MH. The effectiveness of acupuncture in treating acute dental pain: a systematic review. Br Dent J 1998; **184**: 443-447.

- (121) ERNST E, PITTLER MH. Yohimbine for erectile dysfunction: a systematic review and meta-analysis of randomized clinical trials. *J Urol* 1998; **159**: 433-436.
- (122) ERNST E, PITTLER MH. Ginkgo biloba for demetia: a systematic review of double-blind, placebo-controlled trials. *Clin Drug Invest* 1999; **17**: 301-308.
- (123) ERNST E, PITTLER MH. Efficacy of ginger for nausea and vomiting: a systematic review of randomized clinical trials. *Br J Anaesth* 2000; **84**: 367-371.
- (124) ERNST E, RESCH KL. Clinical trials of homoeopathy: a re-analysis of a published review. *Forsch Komplementärmed* 1996; **3**: 85-90.
- (125) ERNST E, STEVINSON C. Ginkgo biloba for tinnitus: a review. *Clin Otolaryngol* 1999; **24**: 164-167.
- (126) ERNST E, WHITE AR. Acupuncture as an adjuvant therapy in stroke rehabilitation. *Wien med Wschr* 1996; **146**: 556-558.
- (127) ERNST E, WHITE AR. Acupuncture for back pain. A meta-analysis of randomized controlled trials. *Arch Intern Med* 1998; **158**: 2235-2241.
- (128) ERNST E, WHITE AR. Acupuncture as a treatment for temporomandibular joint dysfunction. A systematic review of randomised trials. *Arch Otolaryngol Head Neck Surg* 2001; **125**: 269-272.
- (129) EVANS JR. Ginkgo biloba extract for age-related macular degeneration (Cochrane Review). The Cochrane Library, issue 1, 2000. Oxford: Update Software.
- (130) EYSENCK HJ. Meta-analysis and its problems. *BMJ* 1994; **309**: 789-792.
- (131) EZZO J, BERMAN B, HADHAZY VA, JADAD AR, LAO L, SINGH BB. Is acupuncture effective for the treatment of chronic pain? A systematic review. *Pain* 2000; **86**: 217-225.
- (132) FEINSTEIN AR. Meta-analysis: Statistical alchemy for the 21st century. *J Clin Epidemiol* 1995; **48**: 71-79.
- (133) FISCHER MV, BEHR A, VONREUMONT J. Acupuncture - a therapeutic concept in the treatment of painful conditions and functional disorders. Report on 971 cases. *Acupunct Electrother Res Int J* 1984; **9**: 11-29.
- (134) FORMISANO R, FALASCHI P, CERBO R ET AL. Nimodipine in migraine: clinical efficacy and endocrinological effects. *Eur J Clin Pharmacol* 1991; **41**: 69-71.
- (135) FORSSMAN B, HENRIKSSON KG, JOHANNSSON V ET AL. Propranolol for migraine prophylaxis. *Headache* 1976; **16**: 238-245.

- (136) FREITAG FG, DIAMOND S, DIAMOND M. A placebo controlled trial of flunarizine in migraine prophylaxis. *Cephalalgia* 1991; **11 (suppl 11)**: 157-158.
- (137) FRENKEN CW, NUIJTEN ST. Flunarizin - ein neues Mittel zur Migräneprophylaxe. Ergebnisse eines Doppelblindvergleichs mit Plazebo. *Fortschr Med* 1984; **102**: 349-351.
- (138) FRENKEN CW, NUIJTEN ST. Flunarizine, a new preventive approach to migraine. A double-blind comparison with placebo. *Clin Neurol Neurosurg* 1984; **86**: 17-20.
- (139) FRIEDE M, WÜSTENBERG P. Johanniskraut zur Therapie von Angstsyndromen bei depressiven Verstimmungen. *Ztschr Phytother* 1998; **19**: 309-317.
- (140) FULLER GN, GUILOFF RJ. Propranolol in acute migraine: a controlled study. *Cephalalgia* 1990; **10**: 229-233.
- (141) FURUKAWA TA, STREINER DL, HORI S. Discrepancies among megatrials. *J Clin Epidemiol* 2000; **53**: 1193-1199.
- (142) GASTER B, HOLROYD J. St John's wort for depression. A systematic review. *Arch Intern Med* 2000; **160**: 152-156.
- (143) GAWEL M, KREEFT J, NELSON R, SIMARD D. Flunarizine is comparable to propranolol in the prophylaxis of migraine with and without aura. *Cephalalgia* 1991; **11 (suppl 11)**: 156.
- (144) GAWEL MJ, KREEFT J, NELSON RF, SIMARD D, ARNOTT WS. Comparison of the efficacy and safety of flunarizine to propranolol in the prophylaxis of migraine. *Can J Neurol Sci* 1992; **19**: 340-345.
- (145) GERBER WD, DIENER HC, KROPP P, NIEDERBERGER U, SCHOLZ E. The efficacy of flunarizine and etilefrinepivalate in migraine: An empirical double-blind and placebo-controlled study. *Cephalalgia* 1989; **9 (suppl 10)**: 450-451.
- (146) GERBER WD, DIENER HC, SCHOLZ E, NIEDERBERGER U. Responders and non-responders to metoprolol, propranolol and nifedipine treatment in migraine prophylaxis: a dose-range study based on time-series analysis. *Cephalalgia* 1991; **11**: 37-45.
- (147) GERBER WD, SCHELLENBERG R, THOM M ET AL. Cyclandelate versus propranolol in the prophylaxis of migraine--a double-blind placebo-controlled study. *Funct Neurol* 1995; **10**: 27-35.
- (148) GILES JT, PALAT CT, CHIEN SH, CHANG ZG, KENNEDY DT. Evaluation of Echinacea for treatment of the common cold. *Pharmacotherapy* 2000; **20**: 690-697.

- (149) GLANVILLE J, LEFEBVRE C. Wichtige Quellen für systematische Reviews. Evidence-Based Medicine, Dt Ausgabe 2000; **4**: 100-101.
- (150) GÖBEL H, PETERSEN-BRAUN M, SOYKA D. Die Prävalenz von Kopfschmerzen in Deutschland. Der Schmerz 1993; **7**: 287-297.
- (151) GRAY GE, GRAY RN, MCCRORY D, PENZIEN D, RAINS J, HASSELBLAD V. Behavioral and physical treatments for migraine headache. <http://www.clinpol.mc.duke.edu> . 1999.
- (152) GREENLAND S. Quality scores are useless and potentially misleading - reply to "Re: A critical look at some popular analytic methods". Am J Epidemiol 1994; **140**: 300-301.
- (153) GROTEMEYER KH, HUSSTEDT IW, SCHLAKE HP. Betablocker vs Placebo bei vasomotorischem Kopfschmerz. Eine doppelblinde Cross-over-Studie. Dtsch Med Wochenschr 1987; **112**: 1740-1743.
- (154) GROTEMEYER KH, SCHARAFINSKI HW, SCHLAKE HP, HUSSTEDT IW. Acetylsalicylic acid vs. metoprolol in migraine prophylaxis--a double-blind cross-over study. Headache 1990; **30**: 639-641.
- (155) GROTEMEYER KH, SCHLAKE HP, HUSSTEDT IW. Migräneprophylaxe mit Metoprolol und Flunarizin. Eine Doppelblind-cross-over-Studie. Nervenarzt 1988; **59**: 549-552.
- (156) GROTEMEYER KH, SCHLAKE HP, HUSSTEDT IW. Etilefrine pivalate vs. dihydroergotamine and flunarizine in prophylactic treatment of migraine in patients with low blood pressure - a randomized double-blind study. Cephalalgia 1989; **10 (suppl 10)**: 433-434.
- (157) GROTEMEYER KH, SCHLAKE HP, HUSSTEDT IW, RONCOLATO M. Metoprolol versus flunarizine - a double-blind cross-over study. Cephalalgia 1987; **7 (suppl 6)**: 465-466.
- (158) GUCCIONE G, GENTILE F, PULLARA C, SPARACIA B, LIGUORI A, SPARACIA A. Trattamento dell'emicrania senza aura con agopuntura basata sulla differenziazione delle sindromi vs terapia farmacologica e agopuntura standadizzata. In: Gruppo Italiano di ricerca in agopuntura, Hrsg. La ricerca in agopuntura in Italia. Roma: Edizioni Paracelso, 1996: 175-177.
- (159) HALAMA P. Wirksamkeit des Hypericum-Extraktes LI 160 bei 50 Patienten einer psychiatrischen Fachpraxis. Nervenheilkunde 1991; **10**: 305-307.
- (160) HALAMA P. Wirksamkeit des Johanniskrautextraktes LI 160 bei depressiver Verstimmung. Nervenheilkunde 1991; **10**: 250-253.

- (161) HAMMERSCHLAG R, MORRIS MM. Clinical trials comparing acupuncture with biomedical standard care: a criteria-based evaluation of research design and reporting. *Compl Ther Med* 1997; **5**: 133-140.
- (162) HANSEN E, HANSEN JH. Acupuncture treatment of chronic tension headache - a controlled cross-over trial. *Cephalalgia* 1985; **5**: 137-142.
- (163) HANSEN K, SORENSEN P, OLESEN J. A controlled study of flunarizine in migraine. *Acta Neurol Scand* 1984; **69 (suppl 98)**: 266-267.
- (164) HARRER G, HÜBNER WD, PODZUWEIT H. Wirksamkeit und Verträglichkeit des Hypericum-Präparates LI 160 im Vergleich mit Maprotilin. *Nervenheilkunde* 1993; **12**: 297-301.
- (165) HARRER G, HÜBNER WD, PODZUWEIT H. Effectiveness and tolerance of the hypericum extract LI 160 compared to maprotiline: a multi-center double-blind study. *J Geriatr Psychiatry Neurol* 1994; **7 (suppl 1)**: S24-S28.
- (166) HARRER G, SCHMIDT U, KUHN U. "Alternative" Depressionsbehandlung mit einem Hypericum-Extrakt. *TW Neurologie* 1991; **5**: 710-716.
- (167) HARRER G, SCHMIDT U, KUHN U, BILLER A. Comparison of equivalence between the St. John's wort extract LoHyp-57 and fluoxetine. *Arzneimittelforsch/Drug Res* 1999; **49**: 289-296.
- (168) HARRER G, SOMMER H. Therapie leichter/mittelschwerer Depressionen mit Hypericum. *Münch med Wschr* 1993; **135**: 305-309.
- (169) HARRER G, SOMMER H. Treatment of mild/moderate depressions with Hypericum. *Phytomedicine* 1994; **1**: 3-8.
- (170) HARRIS PE. Acupressure: a review of the literature. *Compl Ther Med* 1997; **5**: 156-161.
- (171) HAUG HU, ROBBEN H. Die Akupunktur als Objekt allgemeinmedizinischer Forschung. *Ztschr Allgemeinmed* 1986; **62**: 607-612.
- (172) HAVANKA-KANNIAINEN H, HOKKANEN E, MYLLYLÄ VV. Long acting propranolol in the prophylaxis of migraine. Comparison of the daily doses of 80 mg and 160 mg. *Cephalalgia* 1988; **28**: 607-611.
- (173) HÄNSGEN KD, VESPER J. Antidepressive Wirksamkeit eines hochdosierten Hypericum-Extraktes. *Münch med Wschr* 1996; **138**: 29-33.

- (174) HÄNSGEN KD, VESPER J, PLOCH M. Multizentrische Doppelblindstudie zur antidepressiven Wirksamkeit des Hypericumextraktes LI 160. *Nervenheilkunde* 1993; **12**: 285-289.
- (175) HÄRING B, HAUNS B, HERMANN C, HÜBNER WD, MAIER-LENZ H, MARSCHNER N. A double-blind, placebo-controlled pilot study of LI 160 in combination with chemotherapy in patients with solid tumors. *Abstracts 2nd International Congress on Phytomedicine*. München: 1996: SL-88.
- (176) HEDGES LV, OLKIN I. *Statistical methods for meta-analysis*. San Diego: Academic Press, 1985.
- (177) HEDMAN C, ANDERSEN AR, ANDERSSON PG ET AL. Symptoms of classic migraine attacks: modifications brought about by metoprolol. *Cephalalgia* 1988; **8**: 279-284.
- (178) HEDMAN C, WINTHER K, KNUDSEN JB. The difference between non-selective and beta 1-selective beta-blockers in their effect on platelet function in migraine patients. *Acta Neurol Scand* 1986; **74**: 475-478.
- (179) HENRY P, BAILLE H, DARTIGUES F, JOGEIX M. Traitement de la maladie par acupuncture: étude contrôlée. *Proceedings Premières rencontres Médecines Alternatives, Bordeaux, 12./13.April 1986*, S. 209-216.
- (180) HERBERG KW. Alternative zu synthetischen Psychopharmaka? *Therapiewoche* 1994; **44**: 704-713.
- (181) HESSE J, MOGELVANG B, SIMONSEN H. Acupuncture versus metoprolol in migraine prophylaxis: a randomized trial of trigger point inactivation. *J Intern Med* 1994; **235**: 451-456.
- (182) HEYDENREICH A. Die punktförmige transkutane elektrische Nervenstimulation bei funktionellen vertebrogenen Störungen und bei der Migräne. *Ztschr ges Inn Med* 1988; **43**: 651-653.
- (183) HEYDENREICH A. Die punktförmige transkutane elektrische Nervenstimulation in der Migränetherapie. *Psychiat Neurol med Psychol* 1988; **40**: 717-723.
- (184) HEYDENREICH A. Punktförmige transkutane elektrische Nervenstimulation (PuTENS) im einfachen Placebovergleich bei der Migräne. (Eine prospektive randomisierte Studie). *Ztschr ärztl Fortbild* 1989; **83**: 881-883.
- (185) HILL C, DOYON F. Review of randomized trials of homoeopathy. *Rev Epidem Santé Publ* 1990; **38**: 139-147.

- (186) HOFFMANN J, KÜHL ED. Therapie von depressiven Zuständen mit Hypericin. Z Allgemeinmed 1979; **55**: 776-782.
- (187) HOLDORFF B, SINN M, ROTH G. Propranolol in der Migräneprophylaxe. Eine Doppelblindstudie. Med Klin 1977; **72**: 1115-1118.
- (188) HOLROYD KA, FRANCE JL, CORDINGLEY GE ET AL. Enhancing the effectiveness of relaxation-thermal biofeedback training with propranolol hydrochloride. J Consult Clin Psychol 1995; **63**: 327-330.
- (189) HOLROYD KA, PENZIEN D. Propranolol in the management of recurrent migraine: a meta-analytic review. Headache 1991; **31**: 333-340.
- (190) HOPAYIAN K. The need for caution in interpreting high quality systematic reviews. BMJ 2001; **323**: 681-684.
- (191) HOPFENMÜLLER W. Nachweis der therapeutischen Wirksamkeit eines Ginkgo biloba-Spezialextraktes - Meta-Analyse von 11 klinischen Studien mit Patienten mit Hirnleistungsstörungen im Alter. Arzneim-Forsch/Drug Res 1994; **44(II)**, 1005-1013.
- (192) HOTOPF M. REVIEW: St John's wort is more effective than placebo for treating depressive disorders. Evidence-Based Mental Health 1999; **2**: 49.
- (193) HOTTENROTT K, SOMMER HM, LEHRL S, HAUER H. Der Einfluss von Vitamin E und Johanniskraut-Trockenextrakt auf die Ausdauerleistungsfähigkeit von Wettkampfsportlern. Eine placebo-kontrollierte Doppelblindstudie mit Langstreckenläufern und Triathleten. Dtsche Ztschr Sportmed 1997; **48**: 22-27.
- (194) HU HH, CUNG C, LIU TJ ET AL. A randomized controlled trial on the treatment for acute partial ischemic stroke with acupuncture. Neuroepidemiol 1993; **12**: 106-113.
- (195) HÜBNER WD, LANDE S, PODZUWEIT H. Behandlung larvierter Depressionen mit Johanniskraut. Nervenheilkunde 1993; **12**: 278-280.
- (196) HÜBNER WD, LANDE S, PODZUWEIT H. Hypericum treatment of mild depression with somatic symptoms. J Geriatr Psychiatry Neurol 1994; **7 (suppl. 1)**: S12-S15.
- (197) INTERNATIONAL HEADACHE SOCIETY COMMITTEE ON CLINICAL TRIALS IN MIGRAINE. Guideline for controlled trials of drugs in migraine. First edition. Cephalalgia 1991; **11**: 1-12.
- (198) INTERNATIONAL HEADACHE SOCIETY CLINICAL TRIALS SUBCOMMITTEE. Guideline for controlled trials of drugs in migraine: second edition. Cephalalgia 2000; **20**: 765-786.

- (199) IOANNIDIS JPA, Haidich AB, Pappa M ET AL. Comparison of evidence of treatment effects in randomized and nonrandomized studies. JAMA 2001; **286**: 821-830.
- (200) JACOBS JWG, RASKER JJ, VANRIEL PLCM, GRIEBNAU FWJ, VANDEPUTTE LBA. Alternatieve behandelingswijzen bij reumatische aandoeningen; een literatuuronderzoek. Ned Tijdschr Geneesk 1991; **135**: 317-322.
- (201) JADAD AR, BOOKER L, GAULD M ET AL. The treatment of attention-deficit hyperactivity disorder: an annotated bibliography and critical appraisal of published systematic reviews and metaanalyses. Can J Psychiatr 1999; **44**: 1025-1035.
- (202) JADAD AR, COOK DJ, BROWMAN GP. A guide to interpret discordant systematic reviews. CMAJ 1997; **156**: 1411-1416.
- (203) JADAD AR, COOK DJ, JONES A ET AL. Methodology and reports of systematic reviews and meta-analyses. A comparison of Cochrane reviews with articles published in paper-based journals. JAMA 1998; **280**: 278-280.
- (204) JADAD AR, MCQUAY HJ. Meta-Analyses to evaluate analgesic interventions: a systematic qualitative review of their methodology. J Clin Epidemiol 1996; **49**: 235-243.
- (205) JADAD AR, MOHER M, BROWMAN GP ET AL. Systematic reviews and meta-analyses on treatment of asthma - a critical evaluation. BMJ 2000; **320**: 537-540.
- (206) JADAD AR, MOORE RA, CARROLL D ET AL. Assessing the quality of reports of randomized clinical trials: is blinding necessary? Contr Clin Trials 1996; **17**: 1-12.
- (207) JAILWALA J, IMPERIALE TF, KROENKE K. Pharmacologic treatment of the irritable bowel syndrome: a systematic review of randomized, controlled trials. Ann Intern Med 2000; **133**: 136-147.
- (208) JENSEN LB, JENSEN SB. Effect of acupuncture on tension headache and urinary catecholamine excretion. Scand J Dent Res 1982; **90**: 69-75.
- (209) JENSEN LB, MELSEN B, JENSEN SB. Effect of acupuncture on headache measured by reduction in number of attacks and use of drugs. Scand J Dent Res 1979; **87**: 373-380.
- (210) JENSEN LB, TALLGREN A, TROEST T, JENSEN SB. Effect of acupuncture on myogenic headache. Scand J Dent Res 1977; **85**: 456-470.
- (211) JEPSON RG, KLEIJNEN J, LENG GC. Garlic for lower limb atherosclerosis (Cochrane Review). The Cochrane Library, issue 2, 1997. Oxford: Update Software

- (212) JEPSON RG, MIHALJEVIC L, CRAIG J. Cranberries for the treatment of urinary tract infection (Cochrane Review). The Cochrane Library, issue 4, 1998. Oxford: Update Software.
- (213) JEPSON RG, MIHALJEVIC L, CRAIG J. Cranberries for preventing urinary tract infections (Cochrane Review). The Cochrane Library, issue 1, 2001. Oxford: Update Software.
- (214) JEWELL D, YOUNG G. Interventions for nausea and vomiting in early pregnancy (Cochrane Review). The Cochrane Library, issue 4, 1998. Oxford: Update Software.
- (215) JOHANSSON V, KOSIC S, LINDAHL O, LINDWALL L, TIBBLING L. Effect of acupuncture in tension headache and brainstem reflexes. *Advances in Pain Research and Therapy* 1976; **1**: 839-841.
- (216) JOHNSON D. Neurophysiologische Wirkungen von Hypericum im Doppelblindversuch mit Probanden. *Nervenheilkunde* 1991; **10**: 316-317.
- (217) JOHNSON D, KSCIUK H, WOELK H, SAUERWEIN-GIESE E, FRAUENDORF A. Wirkungen mit Johanniskraut-Extrakt LI 160 im Vergleich mit Maprotilin auf Ruhe-EEG und evozierte Potentiale bei 24 Probanden. *Nervenheilkunde* 1993; **12**: 328-330.
- (218) JOHNSON RH, HORNABROOK RW, LAMBIE DG. Comparison of mefenamic acid and propranolol with placebo in migraine prophylaxis. *Acta Neurol Scand* 1986; **73**: 490-492.
- (219) JONAS WB, LINDE K, RAMIREZ G. HOMEOPATHY AND RHEUMATIC DISEASE. In: Panush RS, Hrsg. *Rheumatic Disease Clinics of North America. Complementary and alternative therapies for rheumatic diseases II*. Orlando: W.B. Saunders Company, 2000: 117-123.
- (220) JOSEY ES, TACKET RL. St. John's wort: a new alternative for depression? *Int J Clin Pharmacol Therap* 1999; **37**: 111-119.
- (221) JOY CB, MUMBY-CROFT R, JOY LA. Polyunsaturated fatty acid (fish or evening primrose oil) for schizophrenia (Cochrane Review). The Cochrane Library, issue 2, 2000. Oxford: Update Software.
- (222) JULIEN J, VALLAT JM, LAGUENY A, DARRIET M. Le traitement prophylactique des migraines par le propranolol. *La Nouvelle Presse Médicale* 1976; **5**: 653.
- (223) JUNNILA SYT. Acupuncture for long-term treatment of headache in a national health center. *Am J Acupunct* 1986; **14**: 351-353.
- (224) JUNNILA SYT. Long-term treatment of chronic pain with acupuncture. Part I. *Acupunct Electro-Ther Res Int J* 1987; **12**: 23-36.

- (225) JÜNI P, WITSCHI A, BLOCH R, EGGER M. The hazards of scoring the quality of clinical trials for meta-analysis. *JAMA* 1999; **282**: 1054-1060.
- (226) KANGASNIEMI P, ANDERSEN AR, ANDERSSON PG ET AL. Classic migraine: effective prophylaxis with metoprolol. *Cephalalgia* 1987; **7**: 231-238.
- (227) KANGASNIEMI P, HEDMAN C. Metoprolol and propranolol in the prophylactic treatment of classical and common migraine. A double-blind study. *Cephalalgia* 1984; **4**: 91-96.
- (228) KANGASNIEMI PJ, NYRKE T, LANG AH, PETERSEN E. Femoxetine - a new 5-HT uptake inhibitor - and propranolol in the prophylactic treatment of migraine. *Acta Neurol Scand* 1983; **68**: 262-267.
- (229) KANIECKI RG. A comparison of divalproex with propranolol and placebo for the prophylaxis of migraine without aura. *Arch Neurol* 1997; **54**: 1141-1145.
- (230) KASS B, NESTVOLD K. Propranolol (Inderal) and clonidine (Catapressan) in the prophylactic treatment of migraine. A comparative trial. *Acta Neurol Scand* 1980; **61**: 351-356.
- (231) KATERNDAHL DA, LAWLER WR. Variability in meta-analytic results concerning the value of cholesterol reduction in coronary heart disease: a meta-analysis. *Am J Epidemiol* 1999; **149**: 429-441.
- (232) KENDRICK T. Prescribing antidepressants in general practice. *BMJ* 1996; **313**: 829-830.
- (233) KIENE H. Klinische Studien zur Misteltherapie der Krebserkrankung. Eine kritische Würdigung. Herdecke: Dissertation, 1989.
- (234) KIENE H. Klinische Studien zur Misteltherapie karzinomatöser Erkrankungen. *Therapeutikon* 1989; **6**: 347-353.
- (235) KIM HL, STRELTZER J, GOEBERT D. St. John's wort for depression: A meta-analysis of well-defined clinical trials. *J Nerv Ment Dis* 1999; **187**: 532-539.
- (236) KIM KC, YOUNT RA. The effect of acupuncture on migraine headache. *Am J Chin Med* 1974; **2**: 407-411.
- (237) KJAERSGARD RM, HOLT LB, BORG L, SOELBERG SP, HANSEN PE. Tolfenamic acid versus propranolol in the prophylactic treatment of migraine. *Acta Neurol Scand* 1994; **89**: 446-450.
- (238) KLAPPER JA. An open label cross-over study of divalproex sodium and propranolol HCl in the prevention of migraine headaches. *Headache Quarterly* 1994; **5**: 50-53.

- (239) KLEIJNEN J. Controlled clinical trials in humans on the effects of garlic supplements. In: Kleijnen J, Hrsg. Food supplements and their efficacy. Maastricht: Rijksuniversiteit Limburg, 1991: 73-82.
- (240) KLEIJNEN J, KNIPSCHILD P. Ginkgo biloba for intermittent claudication and cerebral insufficiency. In: Kleijnen J, Hrsg. Food supplements and their efficacy. Maastricht: Rijksuniversiteit Limburg, 1991: 83-94.
- (241) KLEIJNEN J, KNIPSCHILD P. Ginkgo biloba for cerebral insufficiency. Br J clin Pharm 1992; **34**: 352-358.
- (242) KLEIJNEN J, KNIPSCHILD P. Review article and publication bias. Arznei-Forsch/Drug Res 1992; **42(I)**: 587-591.
- (243) KLEIJNEN J, KNIPSCHILD P. Mistletoe treatment for cancer. Review of controlled trials in humans. Phytomedicine 1994; **1**: 255-260.
- (244) KLEIJNEN J, KNIPSCHILD P, TER RIET G. Garlic, onions and cardiovascular risk factors. A review of the evidence from human experiments with emphasis on commercially available preparations. Br J clin Pharm 1989; **28**: 535-544.
- (245) KLEIJNEN J, KNIPSCHILD P, TER RIET G. Clinical trials of homoeopathy. BMJ 1991; **302**: 316-323.
- (246) KLEIJNEN J, TER RIET G, KNIPSCHILD P. Acupuncture and asthma: a review of controlled trials. Thorax 1991; **46**: 799-802.
- (247) KLEIJNEN J, TER RIET G, KNIPSCHILD P. Evening primrose oil. In: Kleijnen J, Hrsg. Food supplements and their efficacy. Maastricht: Rijksuniversiteit Limburg, 1991: 51-61.
- (248) KNIEBEL R, BURCHARD JM. Zur Therapie depressiver Verstimmungen in der Praxis. Z Allgemeinmed 1988; **64**: 689-696.
- (249) KNIPSCHILD P. Ginseng: pep of nep nep? Pharmaceutisch Weekblad 1988; **123**: 4-11.
- (250) KÖNIG CD. Hypericum perforatum L. (gemeines Johanniskraut) als Therapeutikum bei depressiven Verstimmungszuständen - eine Alternative zu synthetischen Arzneimitteln. Dissertation, Universität Basel, 1993.
- (251) KUBIENA G. Akupunktur bei Migräne - eine Dreijahresstudie mit den Migränepatient(inn)en der Ambulanz des Ludwig-Boltzmann-Institutes für Akupunktur in Wien 1976-1978. Dtsche Ztschr Akup 1985; **28**: 4-14.

- (252) KUBIENA G, NISSEL H, PORENTA G, VEITL M, WESSELY P. Akupunktur bei Migräne. Follow-up Studie. Dtsche Ztschr Akup 1992; **35**: 140-148.
- (253) KUGLER J, SCHMIDT A, GROLL S, WEIDENHAMMER W. Zur Pharmakodynamik eines Hypericum-Extraktes. Untersuchungen bei Patienten mit depressiven Zuständen im Vergleich zu Bromazepam und Placebo. Z Allgemeinmed 1990; **66**: 13-20.
- (254) KUGLER J, WEIDENHAMMER W, SCHMIDT A, GROLL S. Therapie depressiver Zustände. Z Allgemeinmed 1990; **66**: 21-29.
- (255) KUNZ R, FRITSCH L, NEUMAYER HH. Kritische Bewertung von präventiven oder therapeutischen Interventionen. In: Kunz R, Ollenschläger G, Raspe H, Jonitz G, Kolkman FW, Hrsg. Lehrbuch evidenzbasierte Medizin in Klinik und Praxis. Köln: Deutscher Ärzte Verlag, 2000: 120-135.
- (256) KUNZ R, OXMAN AD. The unpredictability paradox: review of empirical comparisons of randomised and non-randomised clinical trials. BMJ 1998; **317**: 1185-1190.
- (257) KURITZKY A, HERING R. Prophylactic treatment of migraine with long acting propranolol - a comparison with placebo. Cephalgia 1987; **7 (suppl 6)**: 457-458.
- (258) LAAKMANN G, DIENEL A, KIESER M. Clinical significance of hyperforin for the efficacy of Hypericum extracts on depressive disorders of different severities. Phytomedicine 1998; **5**: 435-442.
- (259) LAAKMANN G, SCHÜLE C, BAGHAI T, KIESER M. St. John's Wort in mild to moderate depression: the relevance of hyperforin for the clinical efficacy. Pharmacopsychiatry 1998; **31**: 54-59.
- (260) LAITINEN J. Acupuncture for migraine prophylaxis: a prospective clinical study with six month's follow-up. Am J Chin Med 1975; **3**: 271-274.
- (261) LAMSUDIN R, SADJIMIN T. Comparison of the efficacy between flunarizine and nifedipine in the prophylaxis of migraine. Headache 1993; **33**: 335-338.
- (262) LANGOHR HD, GERBER WD, KOLETZKI E, MAYER K, SCHROTH G. Clomipramine and metoprolol in migraine prophylaxis--a double-blind crossover study. Headache 1985; **25**: 107-113.
- (263) LANGOHR HD, REINECKE M, GERBER WD, MANGOLD R. Dihydroergotamin und Flunarizin in der Migräne-Prophylaxe. Eine vergleichende Doppelblind-Studie. Fortschr Med 1988; **106**: 65-70.

- (264) LASTILLA M, ACHILLE P, SAVARESE MA, FRASCELLA G. Verapamil versus flunarizina nella terapia preventiva dell'emicrania. Clin Ter 1990; **132**: 401-404.
- (265) LAUTENSCHLAGER J. Akupunktur bei der Behandlung entzündlich-rheumatischer Erkrankungen. Ztschr Rheumatol 1997; **56**: 8-20.
- (266) LAW M, TANG JL. An analysis of the effectiveness of interventions intended to help people stop smoking. Arch Intern Med 1995; **155**: 1933-1941.
- (267) LAWRENCE RM. Acupuncture for the treatment of headache. Am J Acupunct 1974; **2**: 105-109.
- (268) LAWRENCE V, MULROW C, ACKERMAN R ET AL. Garlic: effects on cardiovascular risks and disease, protective effects against cancer, and clinical adverse effects. Evidence Report/Technology Assessment 20. Agency for Healthcare Research and Quality, Rockville, MD, 2000.
- (269) LAWRENCE V, MULROW C, JACOBS B ET AL. Report on milk thistle: Effects on liver disease and cirrhosis and clinical adverse effects. Evidence Report/Technology Assessment 21. Agency for Healthcare Research and Quality, Rockville, MD, 2000.
- (270) LEE A, DONE ML. The use of nonpharmacologic techniques to prevent postoperative nausea and vomiting: a meta-analysis. Anesth Analg 1999; **88**: 1362-1369.
- (271) LEFEBVRE C, CLARKE MJ. Identifying randomised trials. In: Egger M, Smith GD, Altman D, Hrsg. Systematic reviews in health care: meta-analysis in context. London: BMJ Publishing Group, 2001: 69-86.
- (272) LEHMANN V, BANZHAF E, KUNZE E ET AL. Randomisierte klinisch kontrollierte Studie der Effizienz der Akupunktur im Vergleich zur Elektroakupunktur sowie zur Pharmakotherapie mit Propranolol bei Patienten mit häufiger Migräne. Dtsche Ztschr Akup 1991; **34**: 27-30.
- (273) LEHRL S, WOELK H. Ergebnisse von Messungen der kognitiven Leistungsfähigkeit bei Patienten unter der Therapie mit Johanniskraut. Nervenheilkunde 1993; **12**: 281-284.
- (274) LELORIER J, GREGOIRE G, BENHADDAD A, LAPPIERRE J, DERDERIAN F. Discrepancies between meta-analyses and subsequent large randomized, controlled trials. N Engl J Med 1997; **337**: 536-542.
- (275) LENHARD L, WAITE PM. Acupuncture in the prophylactic treatment of migraine headaches: pilot study. N Z Med J 1983; **96**: 663-666.

- (276) LENOIR S, DEGENRING FH, SALLER R. A double-blind randomised trial to investigate three different concentrations of a standardised fresh plant extract obtained from the shoot tips of *Hypericum perforatum* L. *Phytomedicine* 1999; **6**: 141-146.
- (277) LETZEL H, SCHOOP W. Gingko-biloba-Extrakt EGb 761 und Pentoxifyllin bei Claudicatio intermittens. Sekundäranalyse zur klinischen Wirksamkeit. *VASA* 1992; **21**: 403-410.
- (278) LINDE K, CLAUSIUS N, RAMIREZ G ET AL. Are the clinical effects of homeopathy placebo effects? A meta-analysis of placebo-controlled trials. *Lancet* 1997; **350**: 834-843.
- (279) LINDE K, HONDRAS M, VICKERS A, TER RIET G, MELCHART D. Systematic reviews of complementary therapies - an annotated bibliography. Part 3: Homeopathy. *BMC Complement Alt Med* 2001; **1**: 4. (verfügbar unter www.biomedcentral.com)
- (280) LINDE K, JOBST K, PANTON J. Acupuncture for the treatment of asthma bronchiale (Cochrane Review). The Cochrane Library, issue 1, 1997. Oxford: Update Software.
- (281) LINDE K, JOBST KA. Homeopathy for chronic asthma (Cochrane Review). The Cochrane Library, issue 1, 1998. Oxford: Update Software.
- (282) LINDE K, JONAS WB, MELCHART D, WILLICH S. The methodological quality of randomized controlled trials of homeopathy, herbal medicines and acupuncture. *Int J Epidemiol* 2001; **30**: 526-531.
- (283) LINDE K, MELCHART D. Randomized controlled trials of individualized homeopathy: a state-of-the-art review. *J Alt Complement Med* 1998; **4**: 371-388.
- (284) LINDE K, MELCHART D, WILLICH S. Beschreibung der Therapie in klinischen Studien zur Akupunktur bei chronischen Kopfschmerzen. *Dtsche Ztschr Akup* 2001; **44**: 8-14.
- (285) LINDE K, MULROW CD. St John's wort for depression (Cochrane Review). The Cochrane Library, issue 3, 1998. Oxford: Update Software.
- (286) LINDE K, RAMIREZ G, MULROW CD, PAULS A, WEIDENHAMMER W, MELCHART D. St John's wort for depression - an overview and meta-analysis of randomised clinical trials. *BMJ* 1996; **313**: 253-258.
- (287) LINDE K, SCHOLZ M, RAMIREZ G, CLAUSIUS N, MELCHART D, JONAS WB. Impact of study quality on outcome in placebo-controlled trials of homeopathy. *J Clin Epidemiol* 1999; **52**: 631-636.
- (288) LINDE K, TER RIET G, HONDRAS M, VICKERS A, SALLER R, MELCHART D. Systematic reviews of complementary therapies - an annotated bibliography. Part 2: Herbal medicine. *BMC Complement Alt Med* 2001; **1**: 5. (verfügbar unter www.biomedcentral.com)

- (289) LINDE K, VICKERS A, HONDRAS M ET AL. Systematic reviews of complementary therapies - an annotated bibliography. Part I: Acupuncture. BMC Complement Alt Med 2001; **1**: 3. (verfügbar unter www.biomedcentral.com)
- (290) LINDE K, WORKU F, STÖR W ET AL. Randomized clinical trials of acupuncture for asthma - a systematic review. Forsch Komplementärmed 1996; **3**: 148-155.
- (291) LIPSEY MW, WILSON DB. The efficacy of psychological, educational, and behavioral treatment - confirmation from meta-analysis. American Psychologist 1993; **48**: 1181-1209.
- (292) LOH L, NATHAN PW, SCHOTT GD, ZILKHA KJ. Acupuncture versus medical treatment for migraine and muscle tension headaches. J Neurol Neurosurg Psychiatry 1984; **47**: 333-337.
- (293) LONGWORTH W, MCCARTHY PW. A review of research on acupuncture for the treatment of lumbar disk protrusions and associated neurological symptomatology. J Alt Complement Med 1997; **3**: 55-76.
- (294) LOUIS P. A double-blind placebo-controlled prophylactic study of flunarizine (Sibelium) in migraine. Headache 1981; **21**: 235-239.
- (295) LOUIS P, SCHOENEN J, HEDMAN C. Metoprolol v. clonidine in the prophylactic treatment of migraine. Cephalalgia 1985; **5**: 159-165.
- (296) LOUIS P, SPIERINGS EL. Comparison of flunarizine (Sibelium) and pizotifen (Sandomigran) in migraine treatment: a double-blind study. Cephalalgia 1982; **2**: 197-203.
- (297) LUCKING CH, OESTREICH W, SCHMIDT R, SOYKA D. Flunarizine vs. propranolol in the prophylaxis of migraine: two double-blind comparative studies in more than 400 patients. Cephalalgia 1988; **8 (suppl 8)**: 21-26.
- (298) LUDIN HP. Flunarizine and propranolol in the treatment of migraine. Headache 1989; **29**: 219-224.
- (299) LUTZ C. Quantitative Meta-Analyse empirischer Ergebnisse der Homöopathieforschung. Universität Freiburg: Diplomarbeit Psychologisches Institut, 1993.
- (300) LÜDTKE R, WIESENAUER M. Eine Metaanalyse der homöopathischen Behandlung der Pollinosis mit Galphimia glauca. Wien klin Wschr 1997; **147**: 323-327.

- (301) LÜDTKE R, WILKENS J. Klinische Wirksamkeitsstudien zu Arnica in homöopathischen Zubereitungen. In: Albrecht H, Frühwald M, Hrsg. Karl und Veronica Carstens-Stiftung, Jahrbuch, Band 5. Essen: KVC Verlag, 1998: 97-112.
- (302) MACDONALD R, ISHANI A, RUTKS I, WILT TJ. A systematic review of Cernilton for the treatment of benign prostata hyperplasia. Br J Urol Int 2000; **85**: 836-841.
- (303) MAISSENBACHER J, SCHMIDT U, SCHENK N. Therapie mit Hypericum bei Angstzuständen. TW Neurol Psychiatr 1995; **9**: 65-70.
- (304) MAISSEN CP, LUDIN HP. Vergleich der Wirksamkeit von 5-Hydroxytryptophan und von Propranolol in der Intervallbehandlung der Migräne. Schweiz med Wschr 1991; **121**: 1585-1590.
- (305) MAJOR P, STANDNES B. Akupunkturbehandling ved migrene. Tidsskr Nor Laegeforen 1983; **103**: 686-689.
- (306) MALVEA BP, GWON N, GRAHAM JR. Propranolol prophylaxis of migraine. Headache 1973; **12**: 163-167.
- (307) MANZONI GC, BONO G, SACQUEGNA T ET AL. Flunarizine in common migraine: Italian cooperative trial. I. Short-term results and responders' definition. Cephalalgia 1985; **5 (suppl 2)**: 149-153.
- (308) MARTINEZ B, KASPER S, RUHRMANN B, MÖLLER HJ. Hypericum in der Behandlung von saisonal abhängigen Depressionen. Nervenheilkunde 1993; **12** : 302-307.
- (309) MARTINEZ LJ. Flunarizine (Sibelium) in the prophylaxis of migraine. An open, long-term, multicenter trial. Cephalalgia 1988; **8 (suppl 8)**: 15-20.
- (310) MATHEW NT. Prophylaxis of migraine and mixed headache. A randomized controlled study. Headache 1981; **21**: 105-109.
- (311) MAXION H, KLEIN HJ. Zur Pathogenese und Behandlung der Migräne. Münch med Wschr 1978; **120**: 315-318.
- (312) MCCRORY D, PENZIEN DB, GRAY RN, HASSELBLAD V. Behavioral and physical treatments for tension-type and cervicogenic headache. Verfügbar unter <http://www.fcer.org>. 2000.
- (313) MCINTOSH HM, OLLARIO P. Artemisin derivatives for treating severe malaria (Cochrane Review). The Cochrane Library, issue 2, 2000. Oxford: Update Software.

- (314) MCINTOSH HM, OLLARIO P. Artemisin derivatives for treating uncomplicated malaria (Cochrane Review). The Cochrane Library, issue 2, 2000. Oxford: Update Software.
- (315) MCKIBBIN LS, DOWNIE R. Treatment of migraine headaches using auricular acupuncture techniques. *Laser Ther* 1993; **5**: 23-28.
- (316) MCPARTLAND JM, PRUITT PL. Medical marijuana and its use by the immunocompromised. *Alternat Ther Health Med* 1997; **3**: 39-45.
- (317) MCPHERSON K, BRITTON AR, WENNBURG JE. Are randomized controlled trials controlled? Patient preferences and unblind trials. *J R Soc Med* 1997; **90**: 652-656.
- (318) MEIER B. The science behind hypericum. *Advances in Therapy* 1999; **16**: 135-147.
- (319) MELCHART D, LINDE K, FISCHER P ET AL. Acupuncture for idiopathic headache (Cochrane Review). The Cochrane Library, issue 1, 2001. Oxford: Update Software.
- (320) MELCHART D, LINDE K, FISCHER P, KAESMAYR J. Echinacea for prevention and treatment of the common cold (Cochrane Review). The Cochrane Library, issue 1, 1999. Oxford: Update Software.
- (321) MELCHART D, LINDE K, FISCHER P ET AL. Acupuncture for recurrent headaches: a systematic review of randomized controlled trials. *Cephalalgia* 1999; **19**: 779-786.
- (322) MELCHART D, LINDE K, WORKU F, BAUER R, WAGNER H. Immunomodulation with Echinacea - a systematic review of controlled clinical trials. *Phytomedicine* 1994; **1**: 245-254.
- (323) MENDENOPOULOS G, MANAFI T, LOGOTHETIS J, BOSTANTZOPOULOU S. Flunarizine in the prevention of classical migraine: a placebo-controlled evaluation. *Cephalalgia* 1985; **5 (suppl 2)**: 135-140.
- (324) MICIELI G, PIAZZA D, SINFORIANI E ET AL. Antimigraine drugs in the management of daily chronic headaches: clinical profiles of responsive patients. *Cephalalgia* 1985; **5 (suppl 2)**: 219-224.
- (325) MIKKELSEN B, PEDERSEN KK, CHRISTIANSEN LV. Prophylactic treatment of migraine with tolfenamic acid, propranolol and placebo. *Acta Neurol Scand* 1986; **73**: 423-427.
- (326) MITSIKOSTAS DD, POLYCHRONIDIS I. Valproate versus flunarizine in migraine prophylaxis: a randomized, double-open, clinical trial. *Funct Neurol* 1997; **12**: 267-276.

- (327) MOHER D, COOK DJ, EASTWOOD S, OLKIN I, RENNIE D, STROUP DF. Improving the quality of reports of meta-analyses of randomised controlled trials: the QUOROM statement. *Lancet* 1999; **354**: 1896-1900.
- (328) MOHER D, FORTIN P, JADAD AR ET AL. Completeness of reporting of trials published in languages other than English: implications for conduct and reporting of systematic reviews. *Lancet* 1996; **347**: 363-366.
- (329) MOHER D, JADAD AR, NICHOL G, PENMAN M, TUGWELL P, WALSH S. Assessing the quality of randomized controlled trials: an annotated bibliography of scales and checklists. *Contr Clin Trials* 1995; **16**: 62-73.
- (330) MOHER D, PHAM B, AUSEJO M, SAENZ A, HOOD S, BARBER GG. Pharmacological management of intermittent claudication: a meta-analysis of randomised trials. *Drugs* 2000; **59**: 1057-1070.
- (331) MOHER D, PHAM B, JONES A ET AL. Does quality of reports or randomised trials affect estimates of intervention efficacy reported in meta-analyses? *Lancet* 1998; **352**: 609-613.
- (332) MOHER D, PHAM B, KLASSEN TP ET AL. What contributions do languages other than English make on the results of meta-analyses? *J Clin Epidemiol* 2000; **53**: 964-972.
- (333) MOHER D, SCHULZ KF, ALTMAN D. The CONSORT statement: revised recommendations for improving the quality of reports of parallel-group randomized trials. *JAMA* 2001; **285**: 1987-1991.
- (334) MONTASTRUC JL, SENARD JM. Medicaments anticalciques et prophylaxie de la migraine. *Pathol Biol Paris* 1992; **40**: 381-388.
- (335) MONTGOMERY SA, HÜBNER WD, GRIGOLEIT HG. Efficacy and tolerability of St. John's wort extract compared with placebo in patients with a mild to moderate depressive disorder. *Phytomedicine* 2000; **7 (suppl II)**: 107.
- (336) MORSE PF, HORROBIN DF, MANKU MS ET AL. Meta-analysis of placebo-controlled studies of the efficacy of Epogam in the treatment of atopic eczema. Relationship between plasma essential fatty acid changes and response. *Br J Dermatol* 1989; **121**: 75-90.
- (337) MULROW C, LAWRENCE V, ACKERMAN R. Garlic: effects on cardiovascular risks and disease, protective effects against cancer, and clinical adverse effects. Evidence Report/Technology Assessment No. 20. Rockville: Agency for Healthcare Research and Quality, 2000.

- (338) n.n. Johanniskraut vs. SSRI - neue Erkenntnisse zur Wirksamkeit und Verträglichkeit. *Nervenheilkunde* 2000 **(1)**: 92-93.
- (339) NADELMANN JW, PHIL M, STEVENS J, SAPER JR. Propranolol in the prophylaxis of migraine. *Headache* 1986; **26**: 175-182.
- (340) NAHRSTEDT A, BUTTERWECK V. Biologically active and other chemical constituents of the herb of hypericum perforatum. *Pharmacopsychiatry* 1997; **30 (suppl 2)**: 129-134.
- (341) NAPPI G, SANDRINI G, SAVOINI G, CAVALLINI A, DE-RYSKY C, MICIELI G. Comparative efficacy of cyclandelate versus flunarizine in the prophylactic treatment of migraine. *Drugs* 1987; **33 (suppl 2)**: 103-109.
- (342) NEIL HAW, SILAGY C, LANCASTER T ET AL. Garlic powder in the treatment of moderate hyperlipidemia: a controlled trial and meta-analysis. *J Roy Coll Pract London* 1996; **30**: 329-334.
- (343) NICOLODI M, DEL-BIANCO PL, SICUTERI F. The way to serotonergic use and abuse in migraine. *Int J Clin Pharmacol Res* 1997; **17**: 79-84.
- (344) NUIJTEN STM. Comparative trial of flunarizine and pizotifen in the prophylaxis of migraine. *J Neurol* 1985; **232 (suppl)**: 219.
- (345) NUTI A, LUCETTI C, PAVESE N, DELL'AGNELLO G, ROSSI G, BONUCCELLI U. Long-term follow-up after flunarizine or nimodipine discontinuation in migraine patients. *Cephalalgia* 1996; **16**: 337-340.
- (346) NYRKE T, KANGASNIEMI P, LANG AH, PETERSEN E. Steady-state visual evoked potentials during migraine prophylaxis by propranolol and femoxetine. *Acta Neurol Scand* 1984; **69**: 9-14.
- (347) OAKLEY-BROWNE M, HATCHER S, CHURCHILL R ET AL. Cochrane Depression, Anxiety and Neurosis Group. The Cochrane Library, issue 4, 2001. Oxford: Update Software.
- (348) OKAZAKI K, SADOVE MS, KIM SI, LEE MH, CHENG D. Ryodoraku therapy for migraine headache. *Am J Chin Med* 1975; **3**: 61-70.
- (349) OKEN BS, STORZBACH DM, KAYE JA. The efficacy of Ginkgo biloba on cognitive function in Alzheimer disease. *Arch Neurol* 1998; **55**: 1409-1415.
- (350) OLERUD B, GUSTAVSSON CL, FURBERG B. Nadolol and propranolol in migraine management. *Headache* 1986; **26**: 490-493.

- (351) OLESEN J. Role of calcium entry blockers in the prophylaxis of migraine. Eur Neurol 1986; **25 (suppl 1)**: 72-79.
- (352) OLKIN I. Invited commentary: "a critical look at some popular meta-analysis methods". Am J Epidemiol 1994; **140**: 297-299.
- (353) OLSSON JE, BEHRING HC, FORSSMAN B ET AL. Metoprolol and propranolol in migraine prophylaxis: a double-blind multicentre study. Acta Neurol Scand 1984; **70**: 160-168.
- (354) OSTERHEIDER M, SCHMIDTKE A, BECKMANN H. Behandlung depressiver Syndrome mit Hypericum (Johanniskraut) - eine placebokontrollierte Doppelblindstudie. Fortschr Neurol Psychiatr 1992; **60 (suppl 2)**: 210-211.
- (355) OXMAN AD, GUYATT GH. Validitation of an index of the quality of review articles. J Clin Epidemiol 1991; **44**: 1271-1278.
- (356) OXMAN AD, GUYATT GH, SINGER J ET AL. Agreement among reviewers of review articles. J Clin Epidemiol 1991; **44**: 91-98.
- (357) PALFERMAN TG, GIBBERD FB, SIMMONDS JP. Prophylactic propranolol in the treatment of headache. Br J clin Pharm 1983; **37**: 28-29.
- (358) PANIJEL M. Die Behandlung mittelschwerer Angstzustände. Randomisierte Doppelblindstudie zum klinischen Wirksamkeitsvergleich eines Phytotherapeutikums mit Diazepam. Therapiewoche 1985; **41**: 4659-4668.
- (359) PARK J, HOPWOOD V, WHITE AR, ERNST E. Effectiveness of acupuncture for stroke: a systematic review. J Neurol 2001; **248**: 558-563.
- (360) PARK J, WHITE AR, ERNST E. Efficacy of acupuncture as a treatment for tinnitus. Arch Otolaryngol Head Neck Surg 2000; **126**: 489-492.
- (361) PATEL MS, GUTZWILLER F, PACCAUD F, MARAZZI A. A meta-analysis of acupuncture for chronic pain. Int J Epidemiol 1989; **18**: 900-906.
- (362) PATERNA S, MARTINO SG, CAMPISI D, CASCIO IN, MARSALA BA. Valutazione degli effetti indotti dall'uso di verapamile, flunarizina, diltiazem, nimodipina e placebo nella profilassi dell'emicrania. Studio a doppio cieco randomizzato cross-over. Clin Ter 1990; **134**: 119-125.
- (363) PENZIEN D, JOHNSON C, CARPENTER D, HOLROYD JACKSON K. Drug vs. behavioral treatment of migraine: Long-acting propranolol vs. home-based self-management training. Headache 1990; **30**: 300.

- (364) PETITTI DB. Meta-analysis, decision analysis and cost-effectiveness analysis. Oxford: Oxford University Press, 1994.
- (365) PHILIPP M, KOHNEN R, HILLER KO. Hypericum extract versus imipramine or placebo in patients with moderate depression: randomised multicentre study of treatment for eight weeks. *BMJ* 1999; **319**: 1534-1539.
- (366) PINI LA, FERRARI A, GUIDETTI G, GALETTI G, BARBIERI L, STERNIERI E. Effectiveness of flunarizine in altering electronystagmographic patterns in migraine patients: a preliminary report. *Int J Clin Pharmacol Res* 1986; **6**: 27-32.
- (367) PINI LA, FERRARI A, GUIDETTI G, GALETTI G, STERNIERI E. Influence of flunarizine on the altered electronystagmographic (ENG) recordings in migraine. *Cephalalgia* 1985; **5** (suppl 2): 173-175.
- (368) PINTOV S, LAHAT E, ALSTEIN M, VOGEL Z, BARG J. Acupuncture and the opioid system: implications in management of migraine. *Pediatr Neurol* 1997; **17**: 129-133.
- (369) PISCITELLI S, BURSTEIN AH, CHAITT D, ALFARO RM, FALLOON J. Indinavir concentrations and St John's wort. *Lancet* 2000; **355**: 547-548.
- (370) PITA E, HIGUERAS A, BOLANOS J, PEREZ N, MUNDO A. Propranolol and migraine. A clinical trial. *Arch Farmacol Toxicol* 1977; **3**: 273-278.
- (371) PITTLER MH, ERNST E. Artichoke leaf extract for serum cholesterol reduction. *Perfusion* 1998; **11**: 338-340.
- (372) PITTLER MH, ERNST E. Horse-chestnut seed extract for chronic venous insufficiency. A criteria-based systematic review. *Arch Dermatol* 1998; **134**: 1356-1360.
- (373) PITTLER MH, ERNST E. Peppermint oil for irritable bowel syndrome: a critical review and metaanalysis. *Am J Gastroenterol* 1998; **93**: 1131-1135.
- (374) PITTLER MH, ERNST E. Artemether for severe malaria: a meta-analysis of randomized clinical trials. *Clin Infect Dis* 1999; **28**: 597-601.
- (375) PITTLER MH, ERNST E. Efficacy of kava extract for treating anxiety: systematic review and meta-analysis. *J Clin Psychopharmacol* 2000; **20**: 84-89.
- (376) PITTLER MH, ERNST E. Ginkgo biloba extract for the treatment of intermittent claudication: a meta-analysis of randomized trials. *Am J Med* 2000; **108**: 276-281.
- (377) PÖNTINEN PJ, SALMELA TM. Acupuncture treatment of migraine. In: Sicuteri F, Hrsg. Headache new vistas. Florenz: Biomedical Press, 1977: 251-256.

- (378) PRADALIER A, SERRATRICE G, COLLARD M ET AL. Double-blind controlled trial of long-acting propranolol in migraine prophylaxis. *Cephalalgia* 1989; **9 (suppl 10)**: 367-368.
- (379) PRADALIER A, SERRATRICE G, COLLARD M ET AL. Long-acting propranolol in migraine prophylaxis: results of a double-blind, placebo-controlled study. *Cephalalgia* 1989; **9**: 247-253.
- (380) PRADALIER A, SERRATRICE G, COLLARD M ET AL. Beta-bloquants et migraine. Efficacité, contre placebo, du propranolol à libération prolongée. *Thérapie* 1990; **45**: 441-445.
- (381) QUANDT J, SCHMIDT U, SCHENK N. Ambulante Behandlung leichter und mittelschwerer depressiver Verstimmungen. *Allgemeinarzt* 1993; **2**: 97-102.
- (382) RASCOL A, MONTASTRUC JL, RASCOL O. Flunarizine versus pizotifen: a double-blind study in the prophylaxis of migraine. *Headache* 1986; **26**: 83-85.
- (383) RASMUSSEN BK, JENSEN R, SCHROLL M, OLESEN J. Epidemiology of headache in a general population - a prevalence study. *J Clin Epidemiol* 1991; **44**: 1147-1157.
- (384) RAVEAU LC, BOUSSER MG. Le metoprolol, nouvel antimigraineux de fond. *Presse Med* 1988; **17**: 1805-1809.
- (385) REH C, LAUX P, SCHENK N. Hypericum-Extrakt bei Depressionen - eine wirksame Alternative. *Therapiewoche* 1992; **42**: 1576-1581.
- (386) RENFREW MJ, LANG S. Do cabbage leaves prevent breast engorgement? (Cochrane Review). The Cochrane Library, issue 4, 1998. Oxford: Update Software.
- (387) RESCH KL, ERNST E. Wirksamkeitsnachweise komplementärer Therapien - Literaturanalyse am Beispiel der Akupunktur. *Fortschr Med* 1995; **113(5)**: 41-47.
- (388) ROSEN JA. Observations on the efficacy of propranolol for the prophylaxis of migraine. *Ann Neurol* 1983; **13**: 92-93.
- (389) ROSTED P. The use of acupuncture in dentistry: a review of the scientific validity of published papers. *Oral Diseases* 1998; **4**: 100-104.
- (390) ROTH C. Literature review and critical analysis on the topic of "in- and detoxication experiments in homoeopathy". *Berl J Res Hom* 1991; **1,2**: 111-117.
- (391) RUSCHITZKA F, MEIER PJ, TURINA M, LÜSCHER T, NOLL G. Acute heart transplant rejection due to Saint John's wort. *Lancet* 2000; **355**: 548-549.

- (392) RYAN RE. Comparative study of nadolol and propranolol in prophylactic treatment of migraine. *Am Heart J* 1984; **108**: 1156-1159.
- (393) SARGENT J, SOLBACH P, DAMASIO H ET AL. A comparison of naproxen sodium to propranolol hydrochloride and a placebo control for the prophylaxis of migraine headache. *Headache* 1985; **25**: 320-324.
- (394) SARTORY G, MULLER B, METSCH J, POTHMANN R. A comparison of psychological and pharmacological treatment of pediatric migraine. *Behav Res Ther* 1998; **36**: 1155-1170.
- (395) SCHELLENBERG R, SCHWARZ A, NIEDERBERGER U ET AL. Langzeitwirkung von Cyclandelat und Propanolol bei Migräne nach Beendigung einer viermonatigen medikamentösen Prophylaxe. *Nervenheilkunde* 1997; **16**: 183-187.
- (396) SCHLICH D, BRAUKMANN F, SCHENK N. Behandlung depressiver Zustände mit Hypericinium. *Psycho* 1987; **13**: 440-447.
- (397) SCHMIDT R, OESTREICH W. Flunarizine in migraine prophylaxis: the clinical experience. *J Cardiovasc Pharmacol* 1991; **18 (suppl 8)**: S21-S26.
- (398) SCHMIDT U, HARRER G, KUHN U, BERGER-DEINERT W, LUTHER D. Wechselwirkungen von Hypericin-Extrakt mit Alkohol. *Nervenheilkunde* 1993; **12**: 314-319.
- (399) SCHMIDT U, SCHENK N, SCHWARZ I, VORBERG G. Zur Therapie depressiver Verstimmungen. *Psycho* 1989; **15**: 665-671.
- (400) SCHMIDT U, SOMMER H. Johanniskraut-Extrakt zur ambulanten Therapie der Depression. *Fortschr Med* 1993; **111**: 339-342.
- (401) SCHNEIDER B. Ginkgo-biloba-Extrakt bei peripheren arteriellen Verschußkrankheiten. Meta-Analyse von kontrollierten klinischen Studien. *Arzneim-Forsch/Drug Res* 1992; **42**: 428-436.
- (402) SCHOLZ E, GERBER WD, DIENER HC, LANGOHR HD, REINECKE M. Dihydroergomtamine vs flunarizine vs nifedipine vs metoprolol vs propranolol in migraine prophylaxis: a comparative study based on time series analysis. In: Clifford Rose F, Hrsg. *Advances in headache research*. London: John Libbey, 1987: 139-145.
- (403) SCHRADER E. Equivalence of St John's wort extract (Ze 117) and fluoxetine: a randomized, controlled study in mild-moderate depression. *Intern Clin Psychopharmacol* 2000; **15**: 61-68.

- (404) SCHRADER E, MEIER B, BRATTSTRÖM A. Hypericum treatment of mild-moderate depression in a placebo-controlled study. A prospective, double-blind, randomized, placebo-controlled, multicentre study. *Human Psychopharmacol* 1998; **13**: 163-169.
- (405) SCHULZ H, JOBERT M. Der Einfluß von Johanniskraut auf das Schlaf-EEG bei älteren Probandinnen. *Nervenheilkunde* 1993; **12**: 323-327.
- (406) SCHULZ KF, CHALMERS I, HAYES RJ, ALTMAN DG. Empirical evidence of bias. *JAMA* 1995; **273**: 408-412.
- (407) SHELTON RC, KELLER MB, GELENBERG A ET AL. Effectiveness of St John's wort in major depression. A randomized controlled trial. *JAMA* 2001; **285**: 1978-1986.
- (408) SHIFMAN AC. The clinical response of 328 private patients to acupuncture therapy. *Am J Chin Med* 1975; **3**, 165-179.
- (409) SHIMELL CJ, FRITZ VU, LEVIEN SL. A comparative trial of flunarizine and propranolol in the prevention of migraine. *S Afr Med J* 1990; **77**: 75-77.
- (410) SILAGY C. Developing a register of randomised controlled trials in primary care. *BMJ* 1993; **306**: 897-900.
- (411) SILAGY C, NEIL A. Garlic as a lipid lowering agent - a meta-analysis. *J Roy Coll Phys* 1994; **28**: 39-45.
- (412) SILAGY C, NEIL HAW. A meta-analysis of the effect of garlic on blood pressure. *J Hypertens* 1994; **12**: 463-468.
- (413) SINDRUP SH, MADSEN C, BACH FW, GRAM LF, JENSEN TS. St. John's wort has no effect on pain in polyneuropathy. *Pain* 2000; **91**: 361-356.
- (414) SMITH GD, EGGER M, PHILLIPS AN. Beyond the grand mean? *BMJ* 1997; **315**: 1610-1614.
- (415) SMITH LA, OLDMAN AD, MCQUAY HJ, MOORE RA. Teasing apart quality and validity in sytematic reviews: an example from acupuncture trials in chronic neck pain. *Pain* 2000; **86**: 119-132.
- (416) SOLD-DARSEFF J, LEYDHECKER W. Akupunktur bei Schmerzen im Kopfbereich sowie bei Blepharospasmus ohne organische Ursache. *Klin Monatsbl Augenheilkd* 1986; **189**: 167-169.
- (417) SOLOMON GD. Comparative efficacy of calcium antagonist drugs in the prophylaxis of migraine. *Headache* 1985; **25**: 368-371.

- (418) SOLOMON GD, SCOTT AFB. Verapamil and propranolol in migraine prophylaxis: A double-blind, cross-over study. *Headache* 1986; **26**: 325.
- (419) SOMMER H, HARRER G. Placebo-kontrollierte Studie zur Wirksamkeit eines Hypericumpräparates bei 105 Patienten mit Depressionen. *Nervenheilkunde* 1993; **12**: 274-277.
- (420) SOMMER H, HARRER G. Placebo-controlled double-blind study examining the effectiveness of a hypericum preparation in 105 mildly depressed patients. *J Geriatr Psychiatry Neurol* 1994; **7 (suppl 1)**: S9-S11.
- (421) SÖRENSEN PS, HANSEN K, OLESEN J. A placebo-controlled, double-blind, cross-over trial of flunarizine in common migraine. *Cephalalgia* 1986; **6**: 7-14.
- (422) SÖRENSEN PS, LARSEN BH, RASMUSSEN MJ ET AL. Flunarizine versus metoprolol in migraine prophylaxis: a double-blind, randomized parallel group study of efficacy and tolerability. *Headache* 1991; **31**: 650-657.
- (423) SOVAK M, KUNZEL M, STERNBACH RA, DALESSION DJ. Mechanism of the biofeedback therapy of migraine: Volitional manipulation of the psychophysiological background. *Headache* 1980; **21**: 89-92.
- (424) SOYKA D, OESTREICH W. Flunarizin versus Propranolol in der Intervallprophylaxe der Migräne - eine multizentrische Doppelblindstudie in 12 Kliniken. *Nervenheilkunde* 1987; **6**, 177-183.
- (425) SOYKA D, OESTREICH W. Therapeutic effectiveness of flunarizine and propranolol in the interval therapy of migraine. *Cephalalgia* 1987; **7 (suppl 6)**: 467-468.
- (426) SOYKA D, OESTREICH W. Flunarizin versus Propranolol in der Intervallbehandlung der Migräne - Multizentrische Doppelblindstudie bei niedergelassenen Allgemeinärzten und Internisten. *Nervenheilkunde* 1990; **9**: 45-49.
- (427) SPIELBERGER E. Johanniskraut-Präparat lindert selbst mittelschwere Depressionen. *Ärztliche Praxis* 1985; **37**: 2546-2547.
- (428) SPIERINGS EL, MESSINGER HB. Flunarizine vs. pizotifen in migraine prophylaxis: a review of comparative studies. *Cephalalgia* 1988; **8 (suppl 8)**: 27-30.
- (429) SPOEREL WE, VERKEY M, LEUNG CY. Acupuncture in chronic pain. *Am J Chin Med* 1976; **4**: 267-279.

- (430) STAFFELDT B, KERB R, BROCKMÖLLER J, PLOCH M, ROOTS I. Pharmakokinetik von Hypericin und Pseudohypericin nach oraler Einnahme des Johanniskraut-Extrakts LI 160 bei gesunden Probanden. *Nervenheilkunde* 1993; **12**: 331-338.
- (431) STANDNES B. The prophylactic effect of timolol versus propranolol and placebo in common migraine: beta-blockers in migraine. *Cephalalgia* 1982; **2**: 165-170.
- (432) STEARDO L, BONUSO S, DI-STASIO E, MARANO E. Selective and non-selective beta-blockers: are both effective in prophylaxis of migraine? A clinical trial versus methysergide. *Acta Neurol Napoli* 1982; **4**: 196-204.
- (433) STEARDO L, MARANO E, BARONE P, DENMAN DW, MONTELEONE P, CARDONE G. Prophylaxis of migraine attacks with a calcium-channel blocker: flunarizine versus methysergide. *J Clin Pharmacol* 1986; **26**: 524-528.
- (434) STEGER W. Depressive Verstimmungen. *Z Allgemeinmed* 1985; **61**: 914-918.
- (435) STEINER TJ, JOSEPH R, HEDMAN C, ROSE FC. Metoprolol in the prophylaxis of migraine: parallel-groups comparison with placebo and dose-ranging follow-up. *Headache* 1988; **28**: 15-23.
- (436) STENSRUD P, SJAASTAD O. Short-term clinical trial of propranolol in racemic form (Inderal), D-propranolol and placebo in migraine. *Acta Neurol Scand* 1976; **53**: 229-232.
- (437) STENSRUD P, SJAASTAD O. Comparative trial of Tenormin (atenolol) and Inderal (propranolol) in migraine. *Ups J Med Sci Suppl* 1980; **31**: 37-40.
- (438) STENSRUD P, SJAASTAD O. Comparative trial of Tenormin (atenolol) and Inderal (propranolol) in migraine. *Headache* 1980; **20**: 204-207.
- (439) STERNE JAC, EGGER M. Funnel plots for detecting bias in meta-analysis: Guidelines on choice of axis. *J Clin Epidemiol* 2001; **54**: 1046-1055.
- (440) STERNE JAC, EGGER M, SMITH GD. Investigating and dealing with publication and other biases. In: Egger M, Smith GD, Altman DG, Hrsg. *Systematic reviews in health care: meta-analysis in context*. London: BMJ Books, 2001: 189-210.
- (441) STEVINSON C, ERNST E. Hypericum for depression. An update of the clinical evidence. *Neuropsychopharmacology* 1999; **9**: 501-505.
- (442) STEVINSON C, ERNST E. Valerian for insomnia: a systematic review of randomized clinical trials. *Sleep Medicine* 2000; **1**: 91-99.

- (443) STEVINSON C, PITTLER MH, ERNST E. Garlic for treating hypercholesterinemia - a meta-analysis of randomized clinical trials. *Ann Intern Med* 2000; **133**: 420-429.
- (444) STUX G. Treatment of migraine with acupuncture and moxibustion. *Acup Med* 1987; **4(2)**: 6.
- (445) SUDILOVSKY A, ELKIND AH, RYAN-RE S, SAPER JR, STERN MA, MEYER JH. Comparative efficacy of nadolol and propranolol in the management of migraine. *Headache* 1987; **27**: 421-426.
- (446) TANERI Z. Flunarizin versus Cyclandelat. *TW Neurologie Psychiatrie* 1998; **12**: 429-430.
- (447) TAVOLA T, GALA C, CONTE G, INVERNIZZI G. Traditional Chinese acupuncture in tension-type headache: a controlled study. *Pain* 1992; **48**: 325-329.
- (448) TAYLOR MA, REILLY D, LLEWELLYN-JONES RH, MCSHARRY C, AITCHISON SH. Randomised controlled trials of homoeopathy versus placebo in perennial allergic rhinitis with overview of four trial series. *BMJ* 2000; **321**: 471-476.
- (449) TER RIET G, KLEIJNEN J, KNIPSCHILD P. Acupunctur en nekpijn/rugpijn. *Huisarts en wetenschap* 1989; **32**: 223-227.
- (450) TER RIET G, KLEIJNEN J, KNIPSCHILD P. Acupunctuur bij migraine en spanningshoofdpijn. *Huisarts en wetenschap* 1989; **32**: 258-263.
- (451) TER RIET G, KLEIJNEN J, KNIPSCHILD P. Acupunctuur en aangezichtspijn. *Huisarts en wetenschap* 1989; **32**: 264-266.
- (452) TER RIET G, KLEIJNEN J, KNIPSCHILD P. Acupunctuur en astma. *Huisarts en wetenschap* 1989; **32**: 267-272.
- (453) TER RIET G, KLEIJNEN J, KNIPSCHILD P. Acupunctuur en reumatoide artritis. *Huisarts en wetenschap* 1989; **32**: 228-229.
- (454) TER RIET G, KLEIJNEN J, KNIPSCHILD P. Acupunctuur en stoppen met roken. *Huisarts en wetenschap* 1989; **32**: 299-303.
- (455) TER RIET G, KLEIJNEN J, KNIPSCHILD P. Acupunctuur en chronische pijn. *Huisarts en wetenschap* 1989; **32**: 230-238.
- (456) TER RIET G, KLEIJNEN J, KNIPSCHILD P. A meta-analysis of studies into the effect of acupuncture in addiction. *Br J Gen Prac* 1990; **40**: 379-382.
- (457) TER RIET G, KLEIJNEN J, KNIPSCHILD P. Acupuncture and chronic pain: a criteria-based meta-analysis. *J Clin Epidemiol* 1990; **43**: 1191-1199.

- (458) TFELT-HANSEN P. Efficacy of beta-blockers in migraine. A critical review. *Cephalalgia* 1986; **6 (suppl 5)**: 15-24.
- (459) TFELT-HANSEN P, STANDNES B, KANGASNEIMI P, HAKKARAINEN H, OLESEN J. Timolol vs propranolol vs placebo in common migraine prophylaxis: a double-blind multicenter trial. *Acta Neurol Scand* 1984; **69**: 1-8.
- (460) THOMAS M, BEHARI M, AHUJA GK. Flunarizine in migraine prophylaxis: an Indian trial. *Headache* 1991; **31**: 613-615.
- (461) THOMPSON SG. Why sources of heterogeneity in meta-analysis should be investigated. *BMJ* 1994; **309**: 1351-1355.
- (462) TOLKSDORF W, KLIMM J, KLIMCZYK K, PENNINGER M. Die Akupunktur zur Behandlung chronischer Kopfschmerzen. Eine prospektive Studie zum Einfluss der Depression auf den Behandlungserfolg. *Dtsche Ztschr Akup* 1988; **31**: 104-109.
- (463) TURK Z, MOSER I. Heilung des chronischen Kopfschmerzes durch Akupunktur. *Erfahrungsheilkunde* 1990; **39**: 724-726.
- (464) TURNER P. Beta-blocking drugs in migraine. *Postgrad Med J* 1984; **60 (suppl 2)**: 51-55.
- (465) VAN TULDER MW, CHERKIN DC, BERMAN B, LAO L, KOES BW. The effectiveness of acupuncture in the management of acute and chronic low back pain. *Spine* 1999; **24**: 1113-1123.
- (466) VAN TULDER MW, CHERKIN DC, BERMAN B, LAO L, KOES BW. Acupuncture for low back pain (Cochrane Review). *The Cochrane Library*, issue 1, 2000. Oxford: Update Software.
- (467) VAN TULDER MW, KOES BW, BOUTER LM. Low back pain in primary care. Effectiveness of diagnostic and therapeutic interventions. Amsterdam: EMGO Institute, 1996.
- (468) VERHAGEN AP, DE VET HCW, DE BIE RA, BOERS M, VAN DEN BRANDT PA. The art of quality assesment of RTCs included in systemic reviews. *J Clin Epidemiol* 2001; **54**: 651-654.
- (469) VERHAGEN AP, DE VET HCW, DE BIE RA, KESSELS AG, BOERS M, BOUTER LM. The Delphi list: a criteria list for quality assessment of randomized clinical trials for conducting systematic reviews developed by Delphi consensus. *J Clin Epidemiol* 1998; **51**: 1235-1241.

- (470) VERNON H, MCDERMAID CS, HAGINO C. Systematic review of randomized clinical trials of complementary/alternative therapies in the treatment of tension-type and cervicogenic headache. *Compl Ther Med* 1999; **7**: 142-155.
- (471) VERSPEELT J, DE-LOCHT P, AMERY WK. Post-marketing cohort study comparing the safety and efficacy of flunarizine and propranolol in the prophylaxis of migraine. *Cephalalgia* 1996; **16**: 328-336.
- (472) VERSPEELT J, DE-LOCHT P, AMERY WK. Postmarketing study of the use of flunarizine in vestibular vertigo and in migraine. *Eur J Clin Pharmacol* 1996; **51**: 15-22.
- (473) VICKERS A, GOYAL N, HARLAND R, REES R. Do certain countries produce only positive results? A systematic review of controlled trials. *Contr Clin Trials* 1998; **19**: 159-166.
- (474) VICKERS AJ. Can acupuncture have specific effects on health? A systematic review of acupuncture antiemesis trials. *J Roy Soc Med* 1996; **89**: 303-311.
- (475) VICKERS AJ, SMITH C. Analysis of the evidence profile of the effectiveness of complementary therapies in asthma: a qualitative survey and systematic review. *Compl Ther Med* 1997; **5**: 202-209.
- (476) VICKERS AJ, SMITH C. Homoeopathic Oscillocoquinum for preventing influenza and influenza-like syndromes (Cochrane Review). The Cochrane Library, issue 1, 2000. Oxford: Update Software.
- (477) VILMING S, STANDNES B, HEDMAN C. Metoprolol and pizotifen in the prophylactic treatment of classical and common migraine. A double-blind investigation. *Cephalalgia* 1985; **5**: 17-23.
- (478) VINCENT CA. A controlled trial of the treatment of migraine by acupuncture. *Clin J Pain* 1989; **5**: 305-312.
- (479) VINCENT CA. The treatment of tension headache by acupuncture: a controlled single case design with time series analysis. *J Psychosom Res* 1990; **34**: 553-561.
- (480) VOGLER BK, ERNST E. Aloe vera: a systematic review of its clinical effectiveness. *Br J Gen Pract* 1999; **49**: 823-828.
- (481) VOGLER BK, PITTLER MH, ERNST E. Feverfew as a preventive treatment for migraine: a systematic review. *Cephalalgia* 1998; **18**: 704-708.
- (482) VOGLER BK, PITTLER MH, ERNST E. The efficacy of ginseng. A systematic review of randomised clinical trials. *Eur J Clin Pharmacol* 1999; **55**: 567-575.

- (483) VOLZ HP. Controlled clinical trials of Hypericum extracts in depressed patients - an overview. *Pharmacopsychiatry* 1997; **30 (suppl 2)**: 72-76.
- (484) VOLZ HP, EBERHARDT R, GRILL G. Wirksamkeit und Verträglichkeit des Johanniskrautextraktes D-0496 bei leichten bis mittelschweren depressiven Episoden. Plazebokontrollierte Doppelblindstudie über 6 Wochen. *Nervenheilkunde* 2000; **19**: 401-405.
- (485) VOLZ HP, LAUX P. Potential treatment for subthreshold and mild depression: a comparison of St. John's wort extracts and fluoxetine. *Comprehensive Psychiatry* 2000; **41 (suppl 1)**: 133-137.
- (486) VORBACH EU, ARNOLDT KH, HÜBNER WD. Efficacy and tolerability of St. John's wort extract LI 160 versus imipramine in patients with severe depressive episodes according to ICD-10. *Pharmacopsychiatry* 1997; **30 (suppl 2)**: 81-85.
- (487) VORBACH EU, HÜBNER WD, ARNOLDT KH. Wirksamkeit und Verträglichkeit des Hypericum-Extraktes LI 160 im Vergleich mit Imipramin. *Nervenheilkunde* 1993; **12**: 290-296.
- (488) VORBACH EU, HÜBNER WD, ARNOLDT KH. Effectiveness and tolerance of the Hypericum extract LI 160 in comparison with imipramine: randomized double-blind study in 135 outpatients. *J Geriatr Psychiatry Neurol* 1994; **7 (suppl 1)**: 19-23.
- (489) WALACH H. Unspezifische Therapie-Effekte. Das Beispiel Homöopathie. Habilitationsschrift, Psychologisches Institut, Albert-Ludwigs-Universität Freiburg, 1997.
- (490) WARNECKE G. Beeinflussung klimakterischer Depressionen. *Z Allgemeinmed* 1986; **62**: 1111-1113.
- (491) WARSHAFSKY S, KAMER RS, SIVAK SL. Effect of garlic on total serum cholesterol. *Ann Intern Med* 1993; **119**: 599-605.
- (492) WEBER RB, REINMUTH OM. The treatment of migraine with propranolol. *Neurology* 1972; **22**: 366-369.
- (493) WEIHMayr T, ERNST E. Die therapeutische Wirksamkeit von Crataegus. *Fortschr Med* 1996; **114**: 27-29.
- (494) WEINSCHÜTZ T. AKUPUNKTUR BEI KOPFSCHMERZEN. Methodische Grundlagen und Ergebnisse klinischer Untersuchungen. *Der Schmerz* 1996; **10**: 149-155.

- (495) WEINSCHÜTZ TK, NIEDERBERGER U. Zum Stellenwert der Akupunktur in der Migränetherapie. Methodische Grundlagen und Ergebnisse zweier kontrollierter, prospektiver klinischer Studien. *Nervenheilkunde* 1995; **14**: 295-301.
- (496) WEIß G, KALLISCHNIGG G. Ginkgo-biloba-Extrakt (EGb 761) - Meta-Analyse von Studien zum Nachweis der therapeutischen Wirksamkeit bei Hirnleistungsstörungen bzw. peripherer arterieller Verschußkrankheit. *Münch med Wschr* 1991; **10**: 138-142.
- (497) WERTH W. Psychotonin M versus Imipramin in der Chriurgie. *Der Kassenarzt* 1989; **15**: 64-68.
- (498) WETTSTEIN A. Cholinesterasehemmer und Ginkgoextrakte - in der Demenztherapie vergleichbar? Vergleich publizierter plazebokontrollierter Wirksamkeitsstudien von mindestens sechsmonatiger Dauer. *Fortschr Med* 1999; **117**: 11-18.
- (499) WHEATLEY D. LI 160, an extract of St. John's wort, versus amitriptyline in mildly to moderately depressed outpatients - a controlled 6-week clinical trial. *Pharmacopsychiatry* 1997; **30 (suppl 2)**: 77-80.
- (500) WHITE A, EDDLESTON C, HARDIE R, RESCH KL, ERNST E. A pilot study of acupuncture for tension headache, using a novel placebo. *Acupunct Med* 1996; **14**: 11-15.
- (501) WHITE A, ERNST E. A trial method for assessing the adequacy of acupuncture treatments. *Altern Ther Health Med* 1997; **4**: 66-71.
- (502) WHITE AR, ERNST E. A systematic review of randomized controlled trials of acupuncture for neck pain. *Rheumatology* 1999; **38**: 143-147.
- (503) WHITE AR, RAMPES H, ERNST E. Acupuncture for smoking cessation (Cochrane Review). The Cochrane Library, issue 4, 1999. Oxford: Update Software.
- (504) WHITE AR, RESCH KL, ERNST E. Smoking cessation with acupuncture? A 'best evidence synthesis'. *Forsch Komplementärmed* 1997; **4**: 102-105.
- (505) WHITE AR, RESCH KL, ERNST E. A meta-analysis of acupuncture techniques for smoking cessation. *Tobacco Control* 1999; **8**: 393-397.
- (506) WIDEROE TE, VIGANDER T. Propranolol in the treatment of migraine. *BMJ* 1974; **2**: 699-701.
- (507) WIENERT V, CLASSEN R, HILLER KO. Zur Frage der Photosensibilisierung von Hypericin in einer Baldrian-Johanniskraut-Kombination - klinisch-experimentelle, plazebokontrollierte Vergleichsstudie. Unpublizierter Bericht 1991.

- (508) WIESENAUER M, LÜDTKE R. A meta-analysis of the homeopathic treatment of pollinosis with *Galphimia glauca*. *Forsch Komplementärmed* 1996; **3**: 85-90.
- (509) WILLIAMS JW, MULROW CD, CHIQUETTE E, NOEL PH, AGUILAR C, CORNELL J. A systematic review of newer pharmacotherapies for depression in adults: evidence report summary. *Ann Intern Med* 2000; **132**: 743-756.
- (510) WILT TJ, ISHANI A, MACDONALD R, STARK G, MULROW C, LAU J. Beta-sitosterols for benign prostata hyperplasia (Cochrane Review). The Cochrane Library, issue 1, 2000. Oxford: Update Software.
- (511) WILT TJ, ISHANI A, MACDONALD R, STARK G, MULROW C, LAU J. Serenoa repens for treatment of benign prostata hyperplasia (Cochrane Review). The Cochrane Library, issue 1, 2000. Oxford: Update Software.
- (512) WILT TJ, ISHANI A, STARK G, MACDONALD R, LAU J, MULROW C. Saw palmetto extracts for treatment of benign prostatic hyperplasia. A systematic review. *JAMA* 1998; **280**: 1604-1609.
- (513) WILT TJ, MACDONALD R, ISHANI A, RUTKS I, STARK G. Cernilton for benign prostata hyperplasia (Cochrane Review). The Cochrane Library, issue 2, 2000. Oxford: Update Software.
- (514) WINDELER J. Intention to treat - Begründung und Bedeutung. *Int Praxis* 1992; **32**: 641-650.
- (515) WINKEL R, KORITSCH HD, PIAYDA H, TRENCKMANN U. Phytopharmakotherapie der komorbiden depressiven Störungen bei Alkoholabhängigkeit - Wirksamkeit des Johanniskrautextraktes LI 160 bei komorbid-depressiv Alkoholkranken. In: Trenckmann U, Hrsg. *Phytopharmaka und seelische Störungen*. Neu-Isenburg: LinguaMed, 2000: 43-51.
- (516) WINKEL R, KORITSCH HD, PIAYDA H, TRENCKMANN U. St. John's wort extract LI 160 in depressive, alcohol addicted patients. *Phytomedicine* 2000; **7 (suppl II)**: 19.
- (517) WINTHER K, HEDMAN C, FLODGAARD H. Platelet P1, P4-Di (adenosine-51) tetraphosphate (AP4A) in migraine patients before and during beta-adrenoceptor blockade. *Eur J Clin Invest* 1990; **20**: 336-338.
- (518) WITTE B, HARRER G, KAPTAN T, PODZUWEIT H, SCHMIDT U. Behandlung depressiver Verstimmungen mit einem hochkonzentrierten Hypericumpräparat -eine multizentrische plazebokontrollierte Doppelblindstudie. *Fortschr Med* 1995; **113**: 404-408.

- (519) WOBER C, WOBER BC, KOCH G, WESSELY P. Long-term results of migraine prophylaxis with flunarizine and beta-blockers. *Cephalalgia* 1991; **11**: 251-256.
- (520) WOELK H. Comparison of St John's wort and imipramine for treating depression: randomised controlled trial. *BMJ* 2000; **321**: 536-539.
- (521) WONG AHC, SMITH M, BOON HS. Herbal remedies in psychiatric practice. *Arch Gen Psychiatry* 1998; **55**: 1033-1044.
- (522) WORZ R, DRILLISCH C. Migräne-Prophylaxe durch einen Kalziumeintrittsblocker. Ergebnisse einer Doppelblindstudie Flunarizin vs. Pizotifen. *Münch med Wschr* 1983; **125**: 711-714.
- (523) WORZ R, REINHARDT BB, FOH M, GROTEMEYER KH, SCHARAFINSKI HW. Migräneprophylaxe durch Bisoprolol. Ergebnisse einer Doppelblindstudie versus Metoprolol. *Fortschr Med* 1992; **110**: 268-272.
- (524) WORZ R, REINHARDT BB, GROTEMEYER KH, FOH M. Bisoprolol and metoprolol in the prophylactic treatment of migraine with and without aura - a randomized double-blind cross-over multicenter study. *Cephalalgia* 1991; **11 (suppl 11)**: 152-153.
- (525) YAMAUCHI N. The results of therapeutic acupuncture in a pain clinic. *Can Anaesth Soc J* 1976; **23**: 196-206.
- (526) YOUNG GL, JEWELL MD. Creams to prevent striae gravidarum (Cochrane Review). The Cochrane Library, issue 4, 1998. Oxford: Update Software.
- (527) YUE QY, BERGQUIST C, GERDÉN B. Safety of St John's wort (*Hypericum perforatum*). *Lancet* 2000; **355**: 576-577.
- (528) ZELLER K. Once daily administration of *Hypericum* extract (STW 3): a convenient treatment. In: Chrubasik S, Roufogalis BD, Hrsg. Herbal medicinal products for the treatment of pain. Lismore: South Cross University Press, 2000: 164-168.
- (529) ZIBIN D. Treating vascular migraine with deep puncture at fecgchi: a report of 240 cases. *Int J Clin Acup* 1994; **5**: 455-458.
- (530) ZIEGLER DK, HURWITZ A, HASSANEIN RS, KODANAZ HA, PRESKORN SH, MASON J. Migraine prophylaxis. A comparison of propranolol and amitriptyline. *Arch Neurol* 1987; **44**: 486-489.
- (531) ZIEGLER DK, HURWITZ A, PRESKORN S, HASSANEIN R, SEIM J. Propranolol and amitriptyline in the prophylaxis of migraine. Pharmacokinetic and therapeutic effects. *Arch Neurol* 1997; **50**: 825-830.

- (532) ZWÖLFER W, KAISER B, CARTELLIERI M, SPACEK A. Behandlung des Spannungskopfschmerzes mit Akupunktur. Dtsche Ztschr Akup 1994; **37**: 129-132.

9 Zusammenfassung

Systematische Übersichtsarbeiten klinischer Studien spielen bei der Beurteilung der Wirksamkeit prophylaktischer und therapeutischer Interventionen eine zentrale Rolle. Übersichtsarbeiten (im folgenden synonym mit der Bezeichnung Review gebraucht) werden dann als systematisch bezeichnet, wenn sie vordefinierte und transparente Methoden bzgl. Literatursuche, Studienselektion und Studienbewertung verwenden. Systematische Übersichtsarbeiten, bei denen die einbezogenen Studien (Primärstudien) in einer integrierenden statistischen Analyse wie eine einzige große Studie ausgewertet (gepoolt) werden, bezeichnet man als Meta-Analysen.

Die vorliegende Monographie präsentiert 1. zwei größere eigene systematische Übersichtsarbeiten als Anwendungsbeispiele (Kapitel 2), 2. drei empirisch-methodische Untersuchungen zum Thema Studienqualität und deren Auswirkung auf das Ergebnis (Kapitel 3) und 3. eine Analyse der vorliegenden systematischen Übersichtsarbeiten in den Bereichen Akupunktur, Phytotherapie und Homöopathie (Kapitel 4). Sie versteht sich als Beitrag zur methodischen Diskussion und Entwicklung von systematischen Übersichtsarbeiten.

Im ersten Anwendungsbeispiel wurde untersucht, ob Propranolol, Metoprolol und Flunarizin bei der prophylaktischen Behandlung der Migräne wirksamer sind als Placebo, ob die einzelnen Substanzen ähnlich wirksam sind bzw. wie wirksam sie im Vergleich zu anderen Medikamenten sind. Eingeschlossen wurden randomisierte klinische Studien an erwachsenen Migränepatienten mit mindestens 4 Wochen Dauer, in denen eine oder mehrere der genannten Substanzen mit Placebo oder einem anderen Medikament verglichen wurden. Die Literatursuche erfolgte in Medline und der Cochrane Library sowie über das Durchsuchen von Literaturverzeichnissen. Informationen zu Patienten, Interventionen und Ergebnissen wurden mit Hilfe eines vorgetesteten Formulars extrahiert. Die Qualität der Studien wurde mit drei Skalen beurteilt. Soweit möglich erfolgten Effektstärkeberechnungen für Therapieerfolg, Attackenhäufigkeit und Zahl der Patienten mit Nebenwirkungen. Eine quantitative Meta-Analyse wurde aufgrund der inkonsistenten und häufig inadäquaten Ergebnispräsentation in den Primärstudien nicht durchgeführt. Insgesamt entsprachen 57 Studien zu Propranolol (mit 25 Vergleichen vs. Placebo und 46 Vergleichen mit anderen Medikamenten), 16 zu Metoprolol (mit 4 Vergleichen vs. Placebo und 17 vs. andere) und 40 zu Flunarizin (mit 10 Vergleichen vs. Placebo und 36 vs. andere) den Einschlusskriterien. Die Qualität der Studien war in vielen Fällen unbefriedigend. Hauptproblem war die unzureichende Beschreibung von Studienabbrüchen und -ausschlüssen bzw. deren Berücksichtigung in der Analyse. Alle drei Substanzen waren in Bezug auf Therapieerfolg und Attackenzahl Placebo überlegen. Langzeiteffekte über 6 Monate hinaus und nach Absetzen können nicht sicher beurteilt werden. Mit wenigen Ausnahmen ergaben sich in Vergleichen mit anderen Substanzen keine signifikanten Unterschiede; die meisten Studien hatten jedoch zu geringe Fallzahlen, um potenziell

relevante Unterschiede ausschließen zu können. Insgesamt kann die Wirksamkeit von Propranolol, Metoprolol und Flunarizin in der Migräneprophylaxe aufgrund der Konsistenz der Ergebnisse trotz der häufig unbefriedigenden Studienqualität als belegt angesehen werden.

Im zweiten Anwendungsbeispiel wurde untersucht, ob Hypericumextrakte (*Hypericum perforatum*, Johanniskraut) bei der Behandlung von Depressionen (a) wirksamer sind als Placebo, (b) ähnlich wirksam sind wie synthetische Antidepressiva und (c) nebenwirkungsärmer sind als synthetische Antidepressiva. Eingeschlossen wurden randomisierte Studien an depressiven Patienten mit einer Therapiedauer von mindestens 4 Wochen, in denen ein Hypericummonoextrakt mit Placebo oder einem anderen Antidepressivum verglichen wurde. Die Literatursuche erfolgte in verschiedenen Datenbanken, über das Prüfen von Literaturverzeichnissen und über Kontakte mit Experten und Herstellern. Hauptzielkriterium für die Beurteilung der Wirksamkeit war der Responderanteil (meist definiert als ein Rückgang um 50% oder auf einen Wert von max. 10 auf der Hamilton Rating Scale for Depression). 33 Studien (23 Vergleiche mit Placebo, 12 mit anderen Antidepressiva) entsprachen den Einschlusskriterien. Im Vergleich zu Placebo erwiesen sich die Hypericumextrakte als deutlich überlegen, allerdings war der Effekt in neueren Studien und in Studien mit schwerer depressiven Patienten geringer ausgeprägt. Im Vergleich zu älteren und neueren Antidepressiva waren Hypericumextrakte ähnlich effektiv; Nebenwirkungen waren jedoch signifikant seltener. Aufgrund der vorliegenden Daten kann die Wirksamkeit von Hypericumextrakten bei leichten bis mittleren Depressionen als nachgewiesen gelten. Die Heterogenität in den Ergebnissen der einzelnen Studien deutet jedoch darauf hin, dass verschiedene Faktoren (möglicherweise Extraktwahl, Dosierung, Patientenkontext) die Effektivität beeinflussen.

Die erste der drei empirisch-methodischen Analysen in Kapitel 3 untersucht die methodische Qualität randomisierter Studien zu Akupunktur, Phytotherapie und Homöopathie. Insgesamt 207 randomisierte Studien wurden einbezogen; diese waren im Rahmen von 5 systematischen Übersichtsarbeiten gesammelt und bewertet worden. Die Beurteilung der methodischen Qualität erfolgte mithilfe einer validierten Skala (Jadadscore) und Einzelitems. Die methodische Qualität der bewerteten Studien war in hohem Maße variabel, die Mehrzahl der Studien hatte relevante Schwächen. Hauptprobleme waren die Beschreibung der Verblindung der Gruppenzuteilung sowie von Studienabbrüchen und –ausschlüssen. Phytotherapiestudien hatten im Mittel bessere Qualität als Homöopathie- und Akupunkturstudien. Publikationen in einer Medline-gelisteten Zeitschrift, in jüngerer Vergangenheit und in englischer Sprache hatten im Mittel bessere Qualität als ältere, nicht in gelisteten Zeitschriften und in anderen Sprachen veröffentlichte Arbeiten. Die Qualitätsscores der komplementärmedizinischen Studien waren im Schnitt ähnlich wie diejenigen von Studien zu konventionellen Interventionen, die in ähnlichen Analysen bewertet worden waren.

In der zweiten empirisch-methodischen Untersuchung wurde geprüft, ob ein Zusammenhang zwischen methodischer Qualität und Studienergebnis in 89 placebokontrollierten Studien zur

Homöopathie besteht. Die Beurteilung der methodischen Qualität erfolgte mit zwei Skalen und Einzelkriterien. Ein potenzieller Zusammenhang zwischen methodischer Qualität wurde mit drei Methoden untersucht: Prüfung eines Zusammenhangs zwischen einzelnen Qualitätskomponenten und Studienergebnis (Komponentenanalyse), Vergleich von Studien, die eine bestimmte Minimumpunktzahl auf den Qualitätsskalen erreichten bzw. nicht erreichten (Mindestscoreanalyse), und konsekutiver Einschluss von Studien in die Meta-Analyse entsprechend den erreichten Qualitätsscores (kumulative Meta-Analyse). Mit allen drei Methoden ergab sich, dass bessere Studien weniger optimistische Ergebnisse erbracht hatten.

In der dritten empirisch-methodischen Untersuchung wurde am Beispiel der klinischen Studien zu Akupunktur bei Kopfschmerzen die zuletzt angesprochene Thematik auch auf nichtkontrollierte Studien ausgeweitet. Es wurde geprüft, inwieweit sich randomisierte und nichtrandomisierte Studien in Bezug auf Patienten, Interventionen, design-unabhängige Qualitätsaspekte und Ergebnisse unterscheiden. 59 Studien (24 randomisierte und 35 nichtrandomisierte) konnten in die Analyse einbezogen werden. Randomisierte und nichtrandomisierte Studien unterschieden sich bzgl. Patienten und Interventionen zum Teil deutlich. Nichtrandomisierte Studien hatten im Schnitt schlechtere Qualität, jedoch gab es unter den randomisierten Studien sehr schlechte und unter den nichtrandomisierten auch gute Studien. Ein Randomisationsdesign und bessere Qualität waren mit weniger positiven Ergebnissen assoziiert. Die nichtrandomisierten Studien nutzten potenzielle Vorteile (bzgl. Fallzahl, Langzeitbeobachtung, Repräsentativität und Untersuchung prognostischer Faktoren) kaum. Der Einbezug nichtrandomisierter Studien in einen systematischen Review zur Evidenz für die Wirksamkeit der Akupunktur bei Kopfschmerzen hätte daher die Schlussfolgerungen kaum beeinflusst.

In Kapitel 4 wurden die bis Frühjahr 2001 verfügbaren systematischen Übersichtsarbeiten zu Akupunktur, Phytotherapie und Homöopathie zusammengestellt. Die Suche erfolgte primär über die Datenbank des Bereichs Komplementärmedizin der Cochrane Collaboration und in Medline. Um als systematisch klassifiziert zu werden, mußte ein Review mindestens eines der folgenden drei Kriterien erfüllen: explizite Beschreibung von Literatursuche und Einschlusskriterien und/oder formale Beurteilung der methodischen Qualität der eingeschlossenen Studien und/oder Beschreibung und Durchführung einer quantitativen Meta-Analyse. 39 Reviews zur Akupunktur, 58 zur Phytotherapie und 18 zur Homöopathie entsprachen den Einschlusskriterien. Positive Schlussfolgerungen waren im Bereich Phytotherapie am häufigsten, bei der Akupunktur am seltensten. Die Qualität der Reviews war variabel, viele hatten jedoch relevante methodische Schwächen. Siebzehn Fragestellungen wurden von mehr als einer systematischen Übersicht bearbeitet. Die Zahl der eingeschlossenen Primärstudien variierte bei 10 von 17 Vergleichen um mehr als 50%. Unterschiede in den Einschlusskriterien und der Literatursuche erschienen als wahrscheinlichste Ursache für die Diskrepanzen, Unterschiede in der methodischen Qualität der

Reviews schienen dagegen eine geringe Rolle zu spielen. Die Schlussfolgerungen waren trotz der Diskrepanzen häufig ähnlich.

Die vorgelegten Anwendungsbeispiele, empirisch-methodischen Untersuchungen und die Zusammenstellung zu Akupunktur, Phytotherapie und Homöopathie demonstrieren Stärken und Schwächen von systematischen Reviews. Trotz der zu erwartenden methodischen Weiterentwicklungen werden derartige Übersichtsarbeiten auch in Zukunft mit Sorgfalt und Zurückhaltung interpretiert werden müssen.

Nennung der Wissenschaftler, die bei einzelnen systematischen Übersichtsarbeiten bzw. Untersuchungen mitgewirkt haben

Für die Mitarbeit bei den Untersuchungen, die dieser Habilitationsschrift zugrundeliegen, bedanke ich mich bei

- Karin Rossnagel (Berlin, Mitwirkung Propranololreview)
- Hans-Joachim Jaster (Berlin, Mitwirkung Propranololreview)
- Ulrike Euler (Berlin, Mitwirkung Metoprololreview)
- Cindy Mulrow (San Antonio, Mitwirkung Hypericumreview)
- Wolfgang Weidenhammer (München, Mitwirkung Hypericumreview)
- Gregor Moegen (München, Literatursuche und –besorgung)
- Michael Scholz (München, Meta-Regressions-Analysen in Kapitel 3)
- Wayne Jonas (Alexandria, Mitwirkung Homöopathie-Meta-Analyse)
- Nicola Clausius (München, Mitwirkung Homöopathie-Meta-Analyse)
- Gil Ramirez (Fort Worth, Statistik Homöopathie-Meta-Analyse)
- Florian Eitel (München, Mitwirkung Homöopathie-Meta-Analyse)
- Patrizia Fischer (Germering, Mitwirkung Akupunkturreview)
- Andrew Vickers (New York, Mitwirkung Review of Reviews)
- Maria Hondras (Warrenville, Mitwirkung Review of Reviews)
- Gerben ter Riet (Amsterdam, Mitwirkung Review of Reviews)
- Johannes Thormählen (München, Mitwirkung Review of Reviews Akupunktur)
- Brian Berman (Baltimore, Mitwirkung Review of Reviews Akupunktur)
- Reinhard Saller (Zürich, Mitwirkung Review of Reviews Phytotherapie)

Danksagung

Mein besonderer Dank gilt

- meiner Frau Vera und meinen Kindern Rebecca und Johannes, die wesentlich wichtiger sind als alles andere!
- meinen Eltern für ihre finanzielle und sonstige Unterstützung
- der Carl und Veronica Carstens-Stiftung (Essen), ohne deren großzügige finanzielle Unterstützung das Habilitationsvorhaben nicht möglich gewesen wäre
- Dr. Claudia Witt (Berlin), die den Weg nach Berlin wesentlich mitebnete und mir dort oft den Rücken freihielt
- Dr. Andrea Streng (München) für die äußerst sorgfältige und kritische Durchsicht des Manuskriptes
- Dr. Dieter Melchart (München) für seine Unterstützung des Vorhabens, seine Supervision und jahrelange konstruktive Auseinandersetzungen
- Prof. Dr. Stefan N. Willich (Berlin) für seine großzügige Förderung, Unterstützung, Hilfe in kritischen Situationen und Supervision

Publikationen zu in der vorliegenden Habilitationsschrift präsentierten Daten

1. Linde K, Scholz M, Melchart D, Willich S. Should systematic reviews include non-randomized and uncontrolled studies? The case of acupuncture for chronic headache. *J Clin Epidemiol* 2002;55:77-85.
2. Linde K, Vickers A, Hondras M, ter Riet G, Thormählen J, Berman B, Melchart D. Systematic reviews of complementary therapies – an annotated bibliography. Part I. Acupuncture. *BioMed Central Complementary and Alternative Medicine* 2001;1:3 (available online at <http://www.biomedcentral.com/>).
3. Linde K, ter Riet G, Hondras M, Vickers A, Saller R, Melchart D. Systematic reviews of complementary therapies – an annotated bibliography. Part 2. Herbal medicine. *BioMed Central Complementary and Alternative Medicine* 2001;1:5 (available online at <http://www.biomedcentral.com/>).
4. Linde K, Hondras M, Vickers A, ter Riet G, Melchart D. Systematic reviews of complementary therapies – an annotated bibliography. Part 3. Homeopathy. *BioMed Central Complementary and Alternative Medicine* 2001;1:4 (available online at <http://www.biomedcentral.com/>).
5. Linde K, Jonas WB, Melchart D, Willich SN. The methodological quality of randomized controlled trials of homeopathy, herbal medicines and acupuncture. *Int J Epidemiol* 2001;30:526-531.
6. Linde K, Rossnagel K, Jaster HJ, Willich SN. Propranolol for migraine prophylaxis (protocol). In: *The Cochrane Library*, Issue 3, 2001. Oxford: Update Software.
7. Linde K, Scholz M, Ramirez G, Clausius N, Melchart S, Jonas WB. Impact of quality on outcome in placebo-controlled trials of homeopathy. *J Clin Epidemiol* 1999;62: 631-636.